

THÈSES
PRÉSENTÉES DEVANT L'UNIVERSITÉ DE NANCY I

par
Jean-Pierre BERTRAN

pour obtenir
LE GRADE DE DOCTEUR ÈS-SCIENCES MATHÉMATIQUES

1^{re} Thèse

OPTIMISATION STOCHASTIQUE
DANS UN ESPACE DE HILBERT
ET ANALYSE FONCTIONNELLE

2^e Thèse

Propositions données par l'Université

OPÉRATEURS DE CONVOLUTION EN DIMENSION FINIE
ET INFINIE

soutenue le 20 mai 1975, devant le Jury

MM. M. DEPAIX

Président

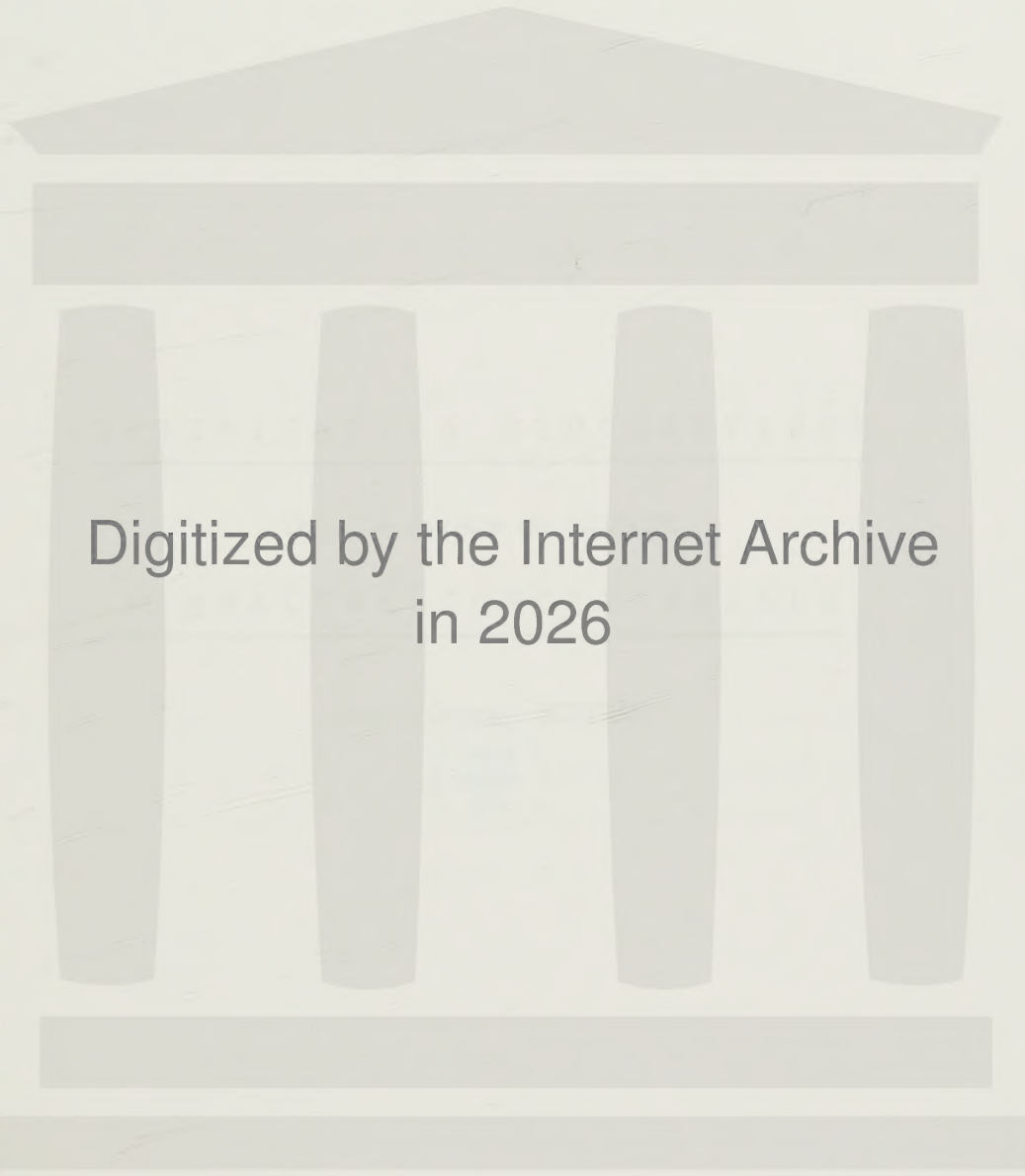
A. BENSOUSSAN

G. CHAVENT

M. LUCIUS

J-P. FERRIER

} Examineurs



Digitized by the Internet Archive
in 2026

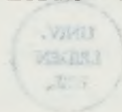
https://archive.org/details/IA41547402_0006

OPTIMISATION STOCHASTIQUE

Dans un espace de Hilbert

et ANALYSE FONCTIONNELLE

Jean-Pierre BERTRAN



UNIVERSITEIT LEIDEN

LIBRARY

LEIDEN



Je suis heureux d'exprimer ici toute ma reconnaissance à Monsieur Alain BENSOUSSAN qui m'a proposé l'objet de ce travail et dont les conseils et les encouragements m'ont permis de le mener à bien.

Monsieur Michel DEPAIX a toujours accepté de me guider dans diverses questions de la théorie des Probabilités. Il a bien voulu accepter de présider ce jury de thèse. Je l'en remercie vivement.

Monsieur Michel LUCIUS et Monsieur Guy CHAVENT m'ont permis d'apprendre la théorie des équations aux dérivées partielles. Ils ont accepté de participer à mon jury, je tiens à les en remercier.

Mes remerciements vont aussi à Monsieur Jean-Pierre FERRIER qui a bien voulu se joindre à mon jury de thèse et me proposer le second sujet.

Enfin, Madame HAMADI, dont j'apprécie depuis plusieurs années la compétence technique, a bien voulu se charger de la réalisation matérielle de ce mémoire. Je la remercie pour le soin et la diligence qu'elle a su y apporter.

TABLE DES MATIERES

Chapitre I

Introduction, outils mathématiques

- I,1 Optimisation stochastique et analyse fonctionnelle
 - I,1,1 Un problème de contrôle pour un système aux dérivées partielles elliptique
 - I,1,2 Un problème de contrôle pour un système aux dérivées partielles parabolique
 - I,1,3 Un problème d'identification de système aux dérivées partielles
- I,2 Le problème de l'optimisation stochastique
 - I,2,1 Enoncé du problème
 - I,2,2 Observation
 - I,2,3 Introduction des fonctionnelles aléatoires linéaires (F.A.L.)
- I,3 Rappels probabilistes
- I,4 Espérance conditionnelle d'une variable aléatoire à valeurs dans Φ
- I,5 Quelques résultats techniques
 - I,5,1 Propriétés de l'opérateur de projection
 - I,5,2 Propriétés des suites de réels
- I,6 Analyse des chapitres suivants

Chapitre II

Optimisation stochastique - méthodes de gradient
l'observation est une variable aléatoire

- II,1 Définition du problème
 - II,1,1 Observation
 - II,1,2 Hypothèse sur f

- II,2 Algorithme de série divergente
 - II,2,1 Algorithme déterministe
 - II,2,2 Algorithme stochastique

- II,3 Algorithme de Gradient stochastique
 - II,3,1 Définition de l'algorithme
 - II,3,2 Convergence de l'algorithme
 - II,3,3 Affaiblissement des hypothèses sur le bruit

- II,4 Comparaison avec les résultats antérieurs
 - II,4,1 Convergence obtenue par la théorie des martingales
 - II,4,2 Convergence obtenue par le lemme de Borel-Cantelli
 - II,4,3 Comparaison entre les différents théorèmes

- II,5 Applications
 - II,5,1 Problème de contrôle dans un système elliptique
 - II,5,2 Problème de contrôle dans un système parabolique
 - II,5,3 Problème d'identification d'un système aux dérivées partielles

- II,6 Méthodes stochastiques de pénalisation
 - II,6,1 Généralités
 - II,6,2 Définition du problème
 - II,6,3 Algorithmes stochastiques, vérification des hypothèses
 - II,6,4 Convergence

Chapitre III

Optimisation stochastique - méthodes de gradient
l'observation est une fonctionnelle aléatoire linéaire

III,1	<u>Problème</u>
III,1,1	<u>Orientation</u>
III,1,2	<u>Le problème stochastique</u>
III,2	<u>Problème approché</u>
III,2,1	<u>Définition de Φ_h</u>
III,2,2	<u>Restriction de l'observation</u>
III,2,3	<u>Problème d'optimisation sur K_h</u>
III,2,4	<u>Hypothèses</u>
III,2,5	<u>Algorithme et convergence</u>
III,3	<u>Convergence de la solution du problème approché</u>
III,4	<u>Exemple d'application</u>

Chapitre V

Méthodes stochastiques de dualité

V,1	<u>Enoncé du problème, données</u>
V,1,1	<u>Données, hypothèses</u>
V,1,2	<u>Problème</u>
V,1,3	<u>Théorèmes d'existence et d'unicité, reformulation du problème dual</u>
V,2	<u>Exemples</u>
V,2,1	<u>Remarques générales</u>
V,2,2	<u>Un problème de contrôle elliptique</u>
V,2,3	<u>Un problème de contrôle parabolique</u>
V,3	<u>Algorithmes stochastiques de recherche de points-selle, généralités</u>
V,3,1	<u>Algorithme d'Uzawa</u>
V,3,2	<u>Algorithme d'Arrow-Hurwicz</u>
V,3,3	<u>Introduction de nouveaux espaces</u>
V,4	<u>Convergence de l'algorithme d'Uzawa stochastique</u>
V,4,1	<u>Convergence de la variable primale</u>
V,4,2	<u>Convergence de la variable duale</u>
V,4,3	<u>Comparaison avec les résultats antérieurs</u>
V,5	<u>Convergence de l'algorithme d'Arrow-Hurwicz stochastique</u>
V,5,1	<u>Convergence de la variable primale</u>
V,5,2	<u>Comparaison avec les résultats antérieurs, problèmes ouverts concernant la convergence de la variable duale</u>

Chapitre VI

Applications numériques

- VI,1 Une application en identification des systèmes aux dérivées partielles
- VI,1,1 Le modèle
- VI,1,2 Application de l'algorithme de gradient (chap. II)
 - VI,1,2,1 Calcul du gradient de la fonctionnelle
 - VI,1,2,2 Remarques sur le contexte théorique
 - VI,1,2,3 Résultats numériques
- VI,1,3 Application de l'algorithme de gradient approché (chap. IV)
 - VI,1,3,1 Calcul du gradient approché
 - VI,1,3,2 Résultats numériques
- VI,2 Un problème de lissage
- VI,2,1 Le modèle
- VI,2,2 Application des théorèmes d'optimisation stochastique
- VI,2,3 Exemple numérique

1

introduction

outils

mathématiques

I.1.- Optimisation stochastique et analyse fonctionnelle

Les problèmes d'optimisation stochastique se présentent naturellement lorsque pour contrôler ou identifier un système, on est en présence d'une observation entachée d'une erreur de mesure.

Le cadre fonctionnel envisagé nous permettra d'étudier les problèmes de contrôle et d'identification dans les systèmes gouvernés par des équations aux dérivées partielles. Dans l'étude des systèmes linéaires avec retard, il permet de rendre compte à la fois des systèmes gouvernés par des équations différentielles ordinaires et des systèmes aux dérivées partielles. Pour préciser les problèmes envisagés exposons dès maintenant quelques exemples importants en théorie des E.D.P.

I.1.1.- Un problème de contrôle pour un système aux dérivées partielles elliptique

Définition des données :

- V : espace de Hilbert sur $\mathbb{R}$
- $a(u,v)$: forme bilinéaire continue sur V , coercive
 $a(u,v) = (Au,v)_{V',V} \quad A \in \mathcal{L}(V,V')$
- $\mathcal{U}$: espace de Hilbert des contrôles
 $B \in \mathcal{L}(\mathcal{U}, V')$
- Soit un système (physique, mécanique, etc...) gouverné par l'opérateur A dont l'état $y(v) \in V$ est solution de $Ay(v) = f + Bv \quad (f \in V')$
 $y(v)$ est défini de façon unique (J.L. Lions [1])
- $\mathcal{H}$: espace de Hilbert des observations
 $\mathcal{G} \in \mathcal{L}(V, \mathcal{H})$: opérateur d'observation
 L'observation est $\mathcal{G}y(v) + \eta$ où η est un bruit
- $N \in \mathcal{L}(\mathcal{U}, \mathcal{U}')$, N hermitien défini positif
 $(Nv,v)_{\mathcal{U}} \geq v \|v\|_{\mathcal{U}}^2, \quad v > 0$
- $\Lambda_{\mathcal{U}}$: isomorphisme canonique de $\mathcal{U}$ sur $\mathcal{U}'$
 $\Lambda_{\mathcal{H}}$: isomorphisme canonique de $\mathcal{H}$ sur $\mathcal{H}'$.

Définition du problème de contrôle :

A tout contrôle $v \in \mathcal{U}$, on associe le coût

$$J(v) = \| \mathcal{E}y(v) - z_d \|_{\mathcal{H}}^2 + (Nv, v)_{\mathcal{U}} \quad \text{où } z_d \text{ est donné dans } \mathcal{H}.$$

Soit $\mathcal{U}_{ad}$: sous-ensemble convexe fermé de $\mathcal{U}$
(ensemble des contrôles admissibles)

Le problème de contrôle est alors :

$$\text{trouver } \inf_{v \in \mathcal{U}_{ad}} J(v).$$

On sait (J.L. Lions [1]) qu'il existe u , unique solution de ce problème.

Le problème stochastique est : trouver un algorithme stochastique convergeant vers u en un certain sens probabiliste au vu de l'observation bruitée $\mathcal{E}y(v) + \eta$.

I.1.2.- Un problème de contrôle pour un système aux dérivées partielles parabolique

Définition des données :

- Soient V, H 2 espaces de Hilbert réels $V \subset H \subset V'$
- $(u, v) \mapsto a(t; u, v)$: famille de formes bilinéaires continues sur V telles que $\forall u, v \in V \quad t \mapsto a(t; u, v)$ soit mesurable sur $]0, T[$
 - . $a(t; u, v) \leq C \|u\|_V \|v\|_V$
 - . $\exists \lambda$ tel que $a(t; u, u) + \lambda \|u\|_H^2 \geq \alpha \|u\|_V^2$, $\alpha > 0$
- alors $a(t; u, v) = (A(t)u, v)_{V', V}$, $A(t)u \in V'$
 $A(t) \in \mathcal{L}(L^2(0, T; V); L^2(0, T; V'))$
- $\mathcal{U}$: espace des contrôles ; $B \in \mathcal{L}(\mathcal{U}, L^2(0, T; V'))$
- f donné $\in L^2(0, T; V')$; y_0 donné $\in H$.
- Soit le système parabolique :

$$\frac{dy(v)}{dt} + A(t) y(v) = f + Bv$$

$$y(v)|_{t=0} = y_0$$

$y(v) \in L^2(0, T; V)$ solution de ce problème est défini de la façon unique (J.L. Lions [1]).

On prend les mêmes notations qu'au §I.1.1 pour définir l'observation, mais ici $\mathcal{E} \in \mathcal{L}(L^2(0, T; V), \mathcal{H})$.

Posant encore $J(v) = \| \mathcal{E}y(v) - z_d \|_{\mathcal{H}}^2 + (Nv, v)_{\mathcal{U}}$.

Le problème de contrôle est

$$\inf_{v \in \mathcal{U}_{ad}} J(v).$$

Et ce problème admet une solution unique (J.L. Lions [1]).

On se pose le même problème stochastique qu'au §I.1.1.

I.1.3.- Un problème d'identification de système aux dérivées partielles

Considérons par exemple un système elliptique :

Aux notations introduites au §I.1.1, on ajoute les suivantes :

$A = A(v)$, $v \in \mathcal{U}$ espace de Hilbert des paramètres.

On considère le système elliptique $A(v)y = f$ ($f \in V'$).

Définition du problème d'identification :

A toute valeur v des paramètres, on associe le coût

$J(v) = \| \mathcal{E}y(v) - z_d \|_{\mathcal{H}}^2$ est le problème d'identification est $\inf_{v \in \mathcal{U}_{ad}} J(v)$.

Le problème stochastique est encore : trouver un algorithme stochastique convergeant vers une solution du problème en un certain sens probabiliste au vu de l'observation bruitée $\mathcal{E}y(v) + \eta$.

I.2.- Le problème de l'optimisation stochastique

I.2.1.- Enoncé du problème

Soit Φ un espace de Hilbert séparable sur $\mathbb{R}^{(1)}$ et $K \subset \Phi$, K convexe fermé non vide.

On se propose de minimiser $f(x)$, $x \in K$ $f : \Phi \longrightarrow \mathbb{R}$.

I.2.2.- Observation

Pour des raisons dues à la nature du problème concret sous-jacent, la seule grandeur accessible à la mesure est pour tout $x \in K$: la variable aléatoire à valeurs dans Φ

$$Y(x) = \nabla f(x) + Z(x), \quad E(Z(x)) = 0$$

(1) Φ est éventuellement de dimension infinie. On pourra donc traiter des applications où Φ est un espace de Hilbert fonctionnel.

ou selon le cas la variable aléatoire à valeurs dans $\mathbb{R}$

$$Y(x) = f(x) + Z(x) \quad , \quad E(Z(x)) = 0 \quad .$$

En pratique, dans une application, on saura calculer à partir de l'observation, une réalisation de $Y(x)$ pour tout x .

Les exemples présentés au §I.1 relèvent tous de ce cadre général grâce au calcul du gradient à l'aide d'un état adjoint (J.L. Lions [1] et G. Chavent [1]) .

I.2.3.- Introduction des fonctionnelles aléatoires linéaires (F.A.L.)

Les processus stochastiques utilisés par les physiciens et les ingénieurs sont rarement de vraies fonctions aléatoires au point de vue mathématique. Ainsi interviennent fréquemment des processus stochastiques du type "bruit blanc".

Le cadre mathématique des fonctionnelles aléatoires linéaires permet de poser de façon mathématiquement correcte les problèmes de statistiques intéressant les physiciens et les ingénieurs (cf. A. Bensoussan [1]) .

Dans les exemples présentés au §I.1, le bruit η ne sera généralement pas une variable aléatoire à valeurs dans un espace de Hilbert Φ , mais une fonctionnelle aléatoire sur Φ . On devra donc généraliser les résultats d'optimisation stochastique au cas où l'observation est la connaissance pour tout $x \in K$ et tout $\varphi \in \Phi$ de la F.A.L. : $Y_{\varphi}(x) = (\nabla f(x), \varphi)_{\Phi} + Z_{\varphi}(x)$ où $Z_{\varphi}(x)$ est une F.A.L. sur Φ' telle que $E(Z_{\varphi}(x)) = 0$.

On rappelle au §I.3 ci-dessous quelques notions sur les F.A.L.

On pourra aussi consulter utilement à ce sujet A. Bensoussan [1] .

I.3.- Rappels probabilistes

Définissons le concept de fonctionnelle aléatoire linéaire.

On note $Mes(\underline{\omega}, \mu, \Phi)$ l'espace des variables aléatoires à valeurs dans Φ .

Définition : Une fonctionnelle aléatoire linéaire sur Φ' est une application linéaire de Φ' dans $Mes(\underline{\omega}, \mu, \mathbb{R})$, c'est-à-dire une famille $X_{\varphi^*}(\omega)$, $\varphi^* \in \Phi'$, de variables aléatoires réelles telles que $X_{\lambda_1 \varphi_1^* + \lambda_2 \varphi_2^*}(\omega) = \lambda_1 X_{\varphi_1^*}(\omega) + \lambda_2 X_{\varphi_2^*}(\omega)$.

La donnée d'une variable aléatoire $X(\omega)$, à valeurs dans Φ définit une fonctionnelle aléatoire sur Φ' par

$$(1.3.1) \quad X_{\varphi^*}(\omega) = \langle X(\omega), \varphi^* \rangle \quad \text{où } \langle ., . \rangle \text{ désigne la dualité } \Phi, \Phi'.$$

Si $X_{\cdot}(\omega)$ est presque sûrement continue de Φ' dans $\mathbb{R}$, elle se met sous la forme (1.3.1) où $X(\omega) \in \text{Mes}(\underline{\omega}, \mu, \Phi)$ car Φ est un Hilbert séparable (cf. Bourbaki [1]).

Dans le cas où Φ est de dimension finie, toute application linéaire est nécessairement continue et la notion de F.A.L. coïncide alors avec celle de variable aléatoire à valeurs dans Φ . En revanche, si Φ est de dimension infinie, la notion de F.A.L. généralise celle de variable aléatoire à valeurs dans Φ .

On trouvera dans A. Bensoussan [1], la définition du bruit blanc en temps (resp. en espace) en tant que F.A.L. sur $\Phi = L^2(0, T)$ (resp. $\Phi = L^2(\Omega)$).

I.4.- Espérance conditionnelle d'une variable aléatoire à valeurs dans Φ .

Définition : Soit $(\underline{\omega}, \mathcal{A}, \mu)$ un espace de probabilité et Φ un espace de Hilbert séparable sur $\mathbb{R}$.

Soit $X(\omega)$ une variable aléatoire à valeurs dans Φ . On suppose que $X \in L^1(\underline{\omega}, \mu, \Phi)$.

Alors l'application $\langle X(\cdot), \varphi^* \rangle : \omega \mapsto \langle X(\omega), \varphi^* \rangle$, $\varphi^* \in \Phi'$ est une variable aléatoire réelle $\in L^1(\underline{\omega}, \mu, \mathbb{R})$ pour tout $\varphi^* \in \Phi'$.

Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. $E^{\mathcal{B}}(\langle X(\cdot), \varphi^* \rangle)$ a un sens.

L'application $\varphi^* \mapsto E^{\mathcal{B}}(\langle X(\omega), \varphi^* \rangle)$ est linéaire.

Puisque $X \in L^1(\underline{\omega}, \mu, \Phi)$, $E^{\mathcal{B}}(\|X(\cdot)\|_{\Phi})$ a un sens.

$$\begin{aligned} |E^{\mathcal{B}}(\langle X(\omega), \varphi^* \rangle)| &\leq E^{\mathcal{B}}(|\langle X(\omega), \varphi^* \rangle|) \leq E^{\mathcal{B}}(\|X(\omega)\|_{\Phi} \|\varphi^*\|_{\Phi'}) = \\ &\|\varphi^*\|_{\Phi'} E^{\mathcal{B}}(\|X(\omega)\|_{\Phi}). \end{aligned}$$

Donc l'application : $\varphi^* \mapsto E^{\mathcal{B}}(\langle X(\omega), \varphi^* \rangle)$ est presque sûrement continue $\Phi' \longrightarrow \mathbb{R}$.

Elle se met donc sous la forme :

$$E^{\mathcal{B}}(\langle X(\omega), \varphi^* \rangle) = \langle U(\omega), \varphi^* \rangle$$

où $U(\cdot) \in \text{Mes}(\underline{\omega}, \mu_{\mathcal{B}}, \Phi)$ ($\mu_{\mathcal{B}}$ = restriction de μ à $\mathcal{B}$)

car Φ est séparable (cf. Bourbaki [1]).

Définition : $U(\cdot)$ sera appelé espérance conditionnelle de $X(\omega)$ par rapport à $\mathcal{B}$ et on note $\underline{U}(\cdot) = E^{\mathcal{B}}(X)(\cdot)$.

Propriétés de $E^{\mathcal{B}}(X)$: Pour simplifier la démonstration et parce que ce sera suffisant pour l'usage fait ultérieurement de l'espérance conditionnelle, limitons-nous à des variables aléatoires $\in L^2(\underline{\omega}, \mu, \Phi)$.

On veut ici généraliser la propriété bien connue pour les variables aléatoires réelles : Si Y est $\mathcal{B}$ -mesurable, $E^{\mathcal{B}}(X.Y) = E^{\mathcal{B}}(X) \cdot Y$.

Proposition 1.4.1.- Si $X, Y \in L^2(\underline{\omega}, \mu, \Phi)$ et si Y est $\mathcal{B}$ -mesurable, alors $E^{\mathcal{B}}(\langle X(\omega), Y(\omega) \rangle) = \langle E^{\mathcal{B}}(X)(\omega), Y(\omega) \rangle$.

Démonstration :

Φ est séparable. Soit donc $\{e_n\}$ une base de Φ , $\langle e_n, e_m \rangle = \delta_{n,m}$

$$X(\omega) = \sum_{n=1}^{\infty} \langle X(\omega), e_n \rangle e_n \quad \langle X(\omega), Y(\omega) \rangle = \sum_{n=1}^{\infty} \langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle$$

$$\text{et } \sum_{n=1}^N \langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle \xrightarrow{N \rightarrow \infty} \langle X(\omega), Y(\omega) \rangle \quad \text{p.s.}$$

$$X(\omega), Y(\omega) \in L^2(\underline{\omega}, \mu, \Phi) \implies \|X(\omega)\|_{\Phi} \|Y(\omega)\|_{\Phi} \in L^1(\underline{\omega}, \mu, \mathcal{R}).$$

Soit Φ_N le sous-espace de Φ engendré par $(e_1, \dots, e_N)$ et p_N l'opérateur de projection sur Φ_N

$$\sum_{n=1}^N \langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle = \langle p_N X(\omega), p_N Y(\omega) \rangle$$

$$\left| \sum_{n=1}^N \langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle \right| \leq \|p_N X(\omega)\|_{\Phi} \|p_N Y(\omega)\|_{\Phi} \leq \|X(\omega)\|_{\Phi} \|Y(\omega)\|_{\Phi} = \xi(\omega) \in L^1(\underline{\omega}, \mu, \mathcal{R}).$$

Grâce au théorème de Fatou-Lebesgue (J. Neveu [1]) :

$$E^{\mathcal{B}}\left(\sum_{n=1}^N \langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle\right) \xrightarrow{N \rightarrow \infty} E^{\mathcal{B}}(\langle X(\omega), Y(\omega) \rangle) \quad \text{presque sûrement}$$

$$E^{\mathcal{B}}\left(\sum_{n=1}^N \langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle\right) = \sum_{n=1}^N E^{\mathcal{B}}(\langle X(\omega), e_n \rangle \langle Y(\omega), e_n \rangle)$$

$$= \sum_{n=1}^N E^{\mathcal{B}}(\langle X(\omega), e_n \rangle) \cdot \langle Y(\omega), e_n \rangle = \sum_{n=1}^N \langle E^{\mathcal{B}}(X)(\omega), e_n \rangle \langle Y(\omega), e_n \rangle$$

(propriété des var. aléat. réelles, Y $\mathcal{B}$ -mes.) (définition de $E^{\mathcal{B}}(X)$)

$$\xrightarrow{N \rightarrow \infty} \langle E^{\mathcal{B}}(X)(\omega), Y(\omega) \rangle.$$

I.5.- Quelques résultats techniques

I.5.1.- Propriétés de l'opérateur de projection

Soit Φ un espace de Hilbert sur $\mathbb{R}$ et K une partie convexe fermée non vide de Φ . On sait qu'il existe un opérateur P de projection sur K .

P n'est pas linéaire si K n'est pas un sous-espace de Φ .

On a cependant les propriétés suivantes (cf. par exemple J. Cea [1]).

Proposition 1.5.1.1.-

$$(1.5.1) \quad (x - Px, z - Px)_{\Phi} \leq 0 \quad \forall z \in K$$

$$(1.5.2) \quad (Px - Py, x - y)_{\Phi} \geq \|Px - Py\|_{\Phi}^2$$

$$(1.5.3) \quad \|Px - Py\|_{\Phi} \leq \|x - y\|_{\Phi}$$

Proposition 1.5.1.2.-

$$\|P(a+u) - Pb\|_{\Phi}^2 \leq \|a-b\|_{\Phi}^2 + 2(u, P(a+u) - Pb)_{\Phi}$$

Démonstration :

$$\begin{aligned} \|P(a+u) - Pb\|_{\Phi}^2 &\leq (a-b+u, P(a+u) - Pb) \\ &\leq \frac{1}{2} \{ \|a-b\|_{\Phi}^2 + \|P(a+u) - Pb\|_{\Phi}^2 \} + (u, P(a+u) - Pb)_{\Phi} \end{aligned}$$

Proposition 1.5.1.3.-

$$\|P(a+u) - Pb\|_{\Phi}^2 \leq \|a-b\|_{\Phi}^2 + 2(u, a-b)_{\Phi} + \|u\|_{\Phi}^2$$

Démonstration :

$$\|P(a+u) - Pb\|_{\Phi}^2 \leq \|a+u-b\|_{\Phi}^2 = \|a-b\|_{\Phi}^2 + 2(u, a-b)_{\Phi} + \|u\|_{\Phi}^2$$

Proposition 1.5.1.4.-

Soit P l'opérateur de projection sur Λ , cône convexe fermé de sommet 0 dans un espace de Hilbert. Alors $(Pu - u, Pu) = 0$.

Démonstration :

D'après (1.5.1) $(Pu - u, z - Pu) \geq 0 \quad \forall z \in \Lambda$

$\iff$ (car Λ est un cône convexe de sommet 0)

$$(Pu - u, z) \geq 0 \quad \forall z \in \Lambda$$

$$(Pu - u, Pu) \geq 0 .$$

Mais, en prenant $z = 0$ dans (1.5.1)

$$(Pu - u, Pu) \leq 0 .$$

I.5.2.- Propriétés des suites de réels

Proposition 1.5.2.1.-

Soit $\{u_n\}$ une suite de réels croissante au sens large telle qu'il existe une suite extraite $\{u_{n_\ell}\}$:

$$u_{n_\ell} \xrightarrow{\ell \rightarrow \infty} u . \text{ Alors } u_n \xrightarrow{n \rightarrow \infty} u .$$

Démonstration : $\exists L(\epsilon)$ tq $\ell \geq L(\epsilon) \implies u - \epsilon \leq u_{n_\ell} \leq u$.

Soit $n \geq n_{L(\epsilon)}$ $\exists \ell_1$ et ℓ_2 tels que $n_{\ell_2} \geq n \geq n_{\ell_1} \geq n_{L(\epsilon)}$.

Alors $u - \epsilon \leq u_{n_{\ell_1}} \leq u_n \leq u_{n_{\ell_2}} \leq u$.

Proposition 1.5.2.2.-

Soit $\{u_n\}$ une suite de réels telle que $u_n - u_{n+1} = v_n + w_n$ où $v_n \geq 0$ et w_n = terme général d'une série convergente et telle que les u_n soient bornés inférieurement. Alors $\sum_n v_n < +\infty$ et u_n a une limite.

Démonstration :

$$u_0 - u_{n+1} = \sum_{k=0}^n v_k + \sum_{k=0}^n w_k \leq u_0 - B , \quad B \text{ tel que } u_n \geq B \quad \forall n$$

$$\sum_{k=0}^{\infty} w_k \text{ converge } \implies \sum_{k=0}^{\infty} v_k < +\infty \quad (\text{somme partielle croissante et majorée})$$

$$\implies u_n \text{ a une limite.}$$

Proposition 1.5.2.3.-

Soient $\{\alpha_n\}$, $\{\rho_n\}$ 2 suites réelles telles que :
 $\rho_n > 0$, $\alpha_n > 0$, $\sum_n \rho_n = +\infty$, $\sum_n \alpha_n \rho_n < +\infty$.
Alors 0 est point d'adhérence de la suite α_n .

Démonstration : Par l'absurde.

Sinon, il existe $\varepsilon > 0$ et N tels que $n \geq N \implies \alpha_n \geq \varepsilon$
 $\implies \varepsilon \rho_n \geq \alpha_n \rho_n$, d'où la contradiction.

I.6.- Analyse des chapitres suivants

Pour traiter d'un point de vue stochastique les problèmes d'optimisation sous contraintes, nous envisagerons 3 grands types de méthodes :

- Les méthodes de gradient avec projection
- Les méthodes de pénalisation
- Les méthodes de dualité.

Chapitre II.-

On minimise $f(x)$, $x \in K$ convexe fermé non vide de Φ dans le cas où l'observation est la variable aléatoire à valeurs dans Φ :

$$\nabla f(x) + Z(x) \text{ , } E(Z(x)) = 0 \text{ .}$$

On étudie d'abord un algorithme stochastique de série divergente. Puis on généralise au cas stochastique les résultats de convergence de l'algorithme de gradient avec projection (A.A. Goldstein [1]). Les hypothèses assurant la convergence seront plus fortes que pour l'algorithme de série divergente et difficiles à vérifier dans la pratique. On démontre donc sous les hypothèses de l'algorithme de série divergente, un résultat de convergence plus faible, mais suffisant dans les applications.

On compare les théorèmes de convergence obtenus avec les résultats publiés antérieurement par d'autres auteurs et on donne plusieurs applications des résultats obtenus.

Pour traiter les contraintes par une autre voie, on envisage enfin les méthodes de pénalisation.

Chapitre III.-

On étend dans ce chapitre, les résultats du chapitre II au cas où l'observation n'est plus une variable aléatoire, mais seulement une fonctionnelle aléatoire linéaire.

Chapitre IV.-

Dans ce chapitre, l'observation ne fournit plus le gradient entaché d'un bruit, mais seulement $f(x)$.

On se limite donc au cas où Φ est de dimension finie et on calcule par différences une valeur approchée du gradient.

Deux algorithmes sont étudiés :

- algorithme de série divergente
- algorithme du gradient avec projection (Goldstein) .

Dans le 2ème cas, on établit d'abord la convergence de l'algorithme déterministe qu'on généralise ensuite au cas d'une observation stochastique. Comme au chapitre II, on établit un résultat plus faible de convergence pour l'algorithme du gradient avec projection, sous les hypothèses de l'algorithme de série divergente.

Enfin on compare les théorèmes de convergence obtenus avec les résultats antérieurs.

Chapitre V.-

On donne d'abord quelques exemples tirés de la théorie du contrôle, de problèmes d'optimisation qui se posent sous la forme $\inf \sup$.

Dans un contexte théorique adapté aux exemples donnés (problèmes de Lagrangien), on étend ensuite au cas d'une observation stochastique les algorithmes d'Uzawa et d'Arrow-Hurwicz bien connus dans le cadre déterministe.

Les résultats de convergence obtenus concernent la variable primale ainsi que la variable duale (pour l'algorithme d'Uzawa).

Chapitre VI.-

On traite complètement deux exemples numériques.

Le premier, en identification d'un opérateur aux dérivées partielles permet de comparer les méthodes des chapitres II et IV.

Le second, un problème de lissage, permet de comparer les différentes méthodes du chapitre II et en particulier de mettre en oeuvre un algorithme "mixte".

2

optimisation stochastique

méthodes de gradient

l'observation est

une variable aléatoire

Chapitre II

Optimisation stochastique - méthodes de gradient
l'observation est une variable aléatoire

II.1.- Définition du problème

Soit Φ un espace de Hilbert séparable sur $\mathbb{R}$
et $f : \Phi \longrightarrow \mathbb{R}$.

On veut minimiser $f(x)$ sur K , convexe fermé non vide contenu dans Φ .

II.1.1.- Observation

L'expression analytique de f est inconnue, mais on connaît, pour tout $x \in K$, la variable aléatoire à valeurs dans Φ : $Y(x) = \nabla f(x) + Z(x)$; $E(Z(x)) = 0$

En pratique, dans une application, on saura calculer, à partir de l'observation, une réalisation de $Y(x)$ pour tout x .

Remarque : Au § II.2.1 ci-dessous on étudie une méthode déterministe dont on prendra ensuite l'analogue stochastique. Dans ce cas, l'observation sera naturellement $\nabla f(x)$.

II.1.2.- Hypothèses sur f

Les hypothèses assurant la convergence sont de 2 types :

- * Hypothèses sur f , nécessaires aussi bien dans le cas déterministe que dans le cas stochastique
- * Hypothèses sur le bruit $Z(x)$ dans le cas stochastique.

Hypothèses sur f :

H_1 : f est bornée inférieurement sur K .

H_2 : f est Frechet-différentiable sur K

H_3 : $\|\nabla f(x) - \nabla f(y)\|_{\Phi} \leq \Lambda \|x-y\|_{\Phi}$

H_4 : $(\nabla f(x) - \nabla f(y), x-y)_{\Phi} \geq \lambda \|x-y\|_{\Phi}^2$, $\lambda \geq 0$.

Remarque 1 : $H_4 \implies f$ est convexe

$$\begin{aligned} f(v) - f(u) &= \int_0^1 (\nabla f(u+\theta(v-u)), v-u)_{\Phi} d\theta = \\ &= (\nabla f(u), v-u)_{\Phi} + \int_0^1 (\nabla f(u+\theta(v-u)) - \nabla f(u), v-u)_{\Phi} d\theta \geq (\nabla f(u), v-u)_{\Phi} \end{aligned}$$

(H_4)

$\iff f$ convexe car f est différentiable.

Remarque 2 : Pour que H_3 soit vérifiée, il suffit que f soit deux fois continûment différentiable et que $\|f''(u)\| \leq \Lambda \quad \forall u \in \Phi$.

Remarque 3 : L'hypothèse H_4 , $\lambda > 0$ jointe à H_3 n'est guère vérifiée que par les fonctionnelles quadratiques. On peut remplacer H_4 par l'hypothèse :

H'_4 : f est strictement convexe.

En effet on démontre : (J.L. Lions, R. Trémolières, R. Glowinski [1], chap. II, lemme 1.1)

Proposition 2.1.2.1.-

Si f est continûment différentiable et strictement convexe, alors quel que soit $M > 0$, il existe une fonction continue $\lambda \mapsto \delta_M(\lambda)$ de $[0, 2M] \rightarrow \mathbb{R}^+$, strictement croissante, $\delta_M(0) = 0$, telle que $(\nabla f(x) - \nabla f(y), x - y) \geq \delta_M(\|x - y\|)$ $\forall x, y$ tels que $\|x\| \leq M, \|y\| \leq M$.

En posant $\delta_M^o(\tau) = \inf_{\substack{\|x - y\| = \tau \\ \|x\|, \|y\| \leq M}} (\nabla f(x) - \nabla f(y), x - y)$

on a $\delta_M(\tau) = 2 \int_0^\tau \delta_M^o(\theta) \frac{d\theta}{\theta}$.

On en déduit : (l'idée d'introduire une telle fonction φ est due à Joseph Rouyer [1]).

Proposition 2.1.2.2.-

Sous les mêmes hypothèses, il existe une fonction $\varphi : t \mapsto \varphi(t)$ de $[0, 2M]$ dans $\mathbb{R}^+$, $\varphi(t) = 0 \iff t = 0$, $\lim_{t \rightarrow t_0} \varphi(t) = 0 \iff t_0 = 0$, telle que $(\nabla f(x) - \nabla f(y), x - y) \geq \|x - y\| \varphi(\|x - y\|)$.

Démonstration :

Posons $\delta_M(\tau) = \tau \varphi(\tau)$. δ_M strictement croissante $\implies \varphi(\tau) \neq 0$ si $\tau \neq 0$.

Montrons que $\varphi(t) \xrightarrow{t \rightarrow 0} 0$, i.e. $\frac{\delta_M(t)}{t} \xrightarrow{t \rightarrow 0} 0$.

La formule de Taylor donne :

$$0 = \delta_M(0) = \delta_M(t) - t \delta'_M(t) + t \varepsilon(t)$$

$$\frac{\delta_M(t)}{t} = \delta'_M(t) - \varepsilon(t) \quad . \quad \text{Or} \quad \delta'_M(t) = \frac{\delta_M^{\circ}(t)}{t} \quad .$$

Donc $\frac{\delta_M(t)}{t} \sim \frac{\delta_M^{\circ}(t)}{t}$ quand $t \longrightarrow 0$.

$$\frac{\delta_M^{\circ}(t)}{t} = \inf_{\substack{\|x-y\| = t \\ \|x\|, \|y\| \leq M}} (\nabla f(x) - \nabla f(y), \frac{x-y}{\|x-y\|}) \leq \|\nabla f(x) - \nabla f(y)\| \leq \Lambda \|x-y\| = \Lambda t$$

grâce à H₃ .

Il en résulte que H'₄ peut remplacer H₄ .

En effet, H'₄ assure que $(\nabla f(x) - \nabla f(y), x-y) \geq 0$.

Donc H₄ est vraie avec $\lambda = 0$, ce qui suffit dans la démonstration des théorèmes (2.2.1.1) (2.2.2.1) (2.3.2.1) où l'on démontre en particulier que les suites X_n sont bornées.

Par conséquent l'hypothèse $\|x\|, \|y\| < M$ faite dans les propositions (2.1.2.1) et (2.1.2.2) sera satisfaite.

Le remplacement de H₄ par H'₄ modifiera parfois les hypothèses à faire sur le pas d'itération ρ_n . Nous l'indiquerons quand ce sera nécessaire, mais pour ne pas alourdir le texte, nous écrirons les démonstrations en utilisant l'hypothèse H₄ .

II.2.- Algorithme de série divergente

On étudiera d'abord un algorithme déterministe qui généralise aux espaces de Hilbert des résultats classiques dans $\mathbb{R}^n$ (cf. B.T. Poljak [1]) , puis on étudie son analogue stochastique.

II.2.1.- Algorithme déterministe

Définition de l'algorithme :

$$\left| \begin{array}{l} x_0 \in K \text{ quelconque} \\ x_{n+1} = P(x_n - \rho_n \nabla f(x_n)) \\ \sum_n \rho_n = +\infty \quad \rho_n \xrightarrow{n \rightarrow \infty} 0 \quad \rho_n > 0 \quad . \end{array} \right.$$

Convergence

Théorème 2.2.1.1.-

$$\text{Sous } \underline{H_1}, \dots, \underline{H_4}, \quad f(x_n) \xrightarrow{n \rightarrow \infty} L = \inf_{x \in K} f(x)$$

Etablissons d'abord :

Lemme 2.2.1.1.-

$$f(x_n) \xrightarrow{n \rightarrow \infty} L'$$

Démonstration : Etablissons que $\{f(x_n)\}$ est une suite décroissante. On aura alors le résultat grâce à $\underline{H_1}$

$$\begin{aligned} f(x_n) - f(x_{n+1}) &= - \int_0^1 (\nabla f(x_n + \theta(x_{n+1} - x_n)), x_{n+1} - x_n)_\Phi \, d\theta \\ &= - (\nabla f(x_n), x_{n+1} - x_n)_\Phi - \int_0^1 d\theta (\nabla f(x_n + \theta(x_{n+1} - x_n)) - \nabla f(x_n), x_{n+1} - x_n)_\Phi \, d\theta \\ \int_0^1 &\leq \int_0^1 \Lambda \theta \|x_{n+1} - x_n\|_\Phi^2 \, d\theta = \frac{\Lambda}{2} \|x_{n+1} - x_n\|_\Phi^2 \end{aligned}$$

grâce à l'inégalité de Schwartz et à $\underline{H_3}$

$$\begin{aligned} f(x_n) - f(x_{n+1}) &\geq - (\nabla f(x_n), x_{n+1} - x_n)_\Phi - \frac{\Lambda}{2} \|x_{n+1} - x_n\|_\Phi^2 \\ - \rho_n (\nabla f(x_n), x_{n+1} - x_n)_\Phi &= (x_n - \rho_n \nabla f(x_n) - x_n, P(x_n - \rho_n \nabla f(x_n)) - P x_n)_\Phi \\ &= \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 \end{aligned}$$

$$\mu^2(\rho_n) \geq 1 \quad \text{d'après (1.5.2)}$$

$$\begin{aligned} f(x_n) - f(x_{n+1}) &\geq - (\nabla f(x_n), x_{n+1} - x_n)_\Phi - \frac{\Lambda}{2} \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 \\ &\geq \left(\frac{1}{\rho_n} - \frac{\Lambda}{2}\right) \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 \end{aligned}$$

$$\rho_n \xrightarrow{n \rightarrow \infty} 0 \quad \text{Donc } n \geq N_0 \implies \frac{1}{\rho_n} - \frac{\Lambda}{2} \geq 0$$

Donc $\{f(x_n)\}$ est décroissante.

Lemme 2.2.1.2.-

$-\rho_n (\nabla f(x_n), x_{n+1} - x_n)_\Phi$
 et $-(\nabla f(x_n), x_{n+1} - x_n)_\Phi$
 sont les termes généraux > 0 de
 séries convergentes

Démonstration :

lemme 2.2.1.1. $\implies$ la série $\{u_n\}$, $u_n = f(x_n) - f(x_{n+1})$
 est convergente à termes ≥ 0 .

$$\text{Donc } \sum_n \left(\frac{1}{\rho_n} - \frac{\Lambda}{2}\right) \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 < +\infty$$

$$n \text{ assez grand } \implies \left(\frac{1}{\rho_n} - \frac{\Lambda}{2}\right) \geq 1 \text{ .}$$

$$\text{Donc } \sum_n \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 < +\infty$$

$$\text{et } \sum_n \frac{\mu^2(\rho_n)}{\rho_n} \|x_{n+1} - x_n\|_\Phi^2 < +\infty$$

$$\text{car } \frac{\mu^2(\rho_n)}{\rho_n} \|x_{n+1} - x_n\|_\Phi^2 = \left(\frac{1}{\rho_n} - \frac{\Lambda}{2}\right) \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 + \frac{\Lambda}{2} \mu^2(\rho_n) \|x_{n+1} - x_n\|_\Phi^2 \text{ .}$$

On en déduit en particulier :

lemme 2.2.1.3.-

$\{x_n\}$ est borné dans Φ

$$\begin{aligned} \text{En effet } \|x_{n+1} - x_n\|_\Phi^2 &= \|P(x_n - \rho_n \nabla f(x_n)) - P(x_n)\|_\Phi^2 \\ &\leq -\rho_n (\nabla f(x_n), x_{n+1} - x_n)_\Phi \end{aligned}$$

$$\implies \sum_n \|x_{n+1} - x_n\|_\Phi^2 < +\infty \text{ .}$$

Démonstration du théorème 2.2.1.1 :

A cause du lemme 2.2.1.1, il suffit de trouver une sous-suite $\{x_{n_k}\}$ telle
 que $f(x_{n_k}) \xrightarrow[k \rightarrow \infty]{} L$.

Soit $\alpha > L$ fixé et $\varepsilon > 0$ tel que $\alpha > \alpha - \varepsilon > L$.

Notons $S_\alpha = \{x \in K \mid f(x) \leq \alpha\}$.

Puisque $K \neq \emptyset$, $S_\alpha \neq \emptyset$.

Hypothèse : (dont on veut démontrer l'absurdité)

$$\exists N_1 (> N_0) \text{ tel que } n > N_1 \implies x_n \notin S_\alpha$$

i.e. $f(x_n) > \alpha$ puisque par définition de x_n , $x_n \in K$.

$$\text{Donc } S_{\alpha-\varepsilon} \subset S_\alpha \subset \{x \mid f(x) \leq f(x_n)\}$$

$$f \text{ convexe} \implies f(x) - f(x_n) \geq (\nabla f(x_n), x - x_n)_\Phi \quad \forall x \in S_{\alpha-\varepsilon}$$

$$\text{ou } -(\nabla f(x_n), x - x_n)_\Phi \geq f(x_n) - f(x) \geq \alpha - (\alpha - \varepsilon) = \varepsilon.$$

Fixons un point $\tilde{x} \in S_{\alpha-\varepsilon}$

$$\begin{aligned} \|x_{n+1} - \tilde{x}\|_\Phi^2 &= \|P(x_n - \rho_n \nabla f(x_n)) - P\tilde{x}\|_\Phi^2 \\ &\leq \|x_n - \tilde{x}\|_\Phi^2 - 2\rho_n (\nabla f(x_n), x_{n+1} - \tilde{x})_\Phi \\ &\quad (\text{prop. 1.5.1.2}) \\ &\leq \|x_n - \tilde{x}\|_\Phi^2 - 2\rho_n (\nabla f(x_n), x_{n+1} - \tilde{x})_\Phi \\ &\quad - 2\rho_n (\nabla f(x_n), \tilde{x} - x_n)_\Phi - 2\rho_n \varepsilon \\ \|x_{n+1} - \tilde{x}\|_\Phi^2 &\leq \|x_n - \tilde{x}\|_\Phi^2 - 2\rho_n (\nabla f(x_n), x_{n+1} - x_n)_\Phi - 2\rho_n \varepsilon. \end{aligned}$$

Itérons cette inégalité de N à $N+m$

$$0 \leq \|x_{N+m+1} - \tilde{x}\|_\Phi^2 \leq \|x_N - \tilde{x}\|_\Phi^2 + 2 \sum_{n=N}^{N+m} -\rho_n (\nabla f(x_n), x_{n+1} - x_n)_\Phi - 2\varepsilon \sum_{n=N}^{N+m} \rho_n$$

Le 2ème terme du 2ème membre est > 0 (lemme 2.2.1.2)
et la série converge

$\sum_n \rho_n = +\infty \implies$ Pour m assez grand, le 2ème membre devient < 0 , d'où l'absurdité.

Par conséquent

$$\forall \alpha > L, \quad \forall N, \quad \exists n_\alpha > N \text{ tel que } x_{n_\alpha} \in S_\alpha.$$

Il existe donc une sous-suite $\{x_{n_k}\}$ telle que $f(x_{n_k}) \xrightarrow[k \rightarrow \infty]{} L$ et par suite, $f(x_n) \xrightarrow[n \rightarrow \infty]{} L$.

Théorème 2.2.1.2.-

Sous les mêmes hypothèses et si de plus $\lambda > 0$,
 $\{z \in K \mid f(z) = L\}$ est réduit à un point, soit u
et $x_n \xrightarrow[n \rightarrow \infty]{} u$ dans Φ fort.

L'unicité de u est classique (cf. J. Cea [1], th. 0.3)

. Montrons d'abord la convergence faible :

K convexe fermé $\implies K$ faiblement fermé

$\{x_n\}$ borné (lemme 2.2.1.3) $\implies \exists$ une sous-suite $\{x_{n_k}\}$ $x_{n_k} \xrightarrow[k \rightarrow \infty]{} v$
dans Φ faible

Comme K est faiblement fermé, $v \in K$

Théorème 2.2.1.1 $\implies f(x_{n_k}) \xrightarrow[k \rightarrow \infty]{} L$

$$L = \lim_{k \rightarrow \infty} f(x_{n_k}) = \underline{\lim}_{k \rightarrow \infty} f(x_{n_k}) \geq f(v)$$

car f est convexe et Frechet-différentiable.

$v \in K \implies f(v) = L \implies v = u$, minimum unique.

Donc la limite est indépendante de la sous-suite choisie et $x_n \xrightarrow[n \rightarrow \infty]{} u$ dans Φ faible.

. Montrons enfin la convergence forte :

$$\begin{aligned} f(x_n) - f(u) &= \int_0^1 (\nabla f(u + \theta(x_n - u)), x_n - u)_{\Phi} d\theta \\ &= (\nabla f(u), x_n - u)_{\Phi} + \int_0^1 (\nabla f(u + \theta(x_n - u)) - \nabla f(u), x_n - u)_{\Phi} d\theta \\ &\geq (\nabla f(u), x_n - u)_{\Phi} + \frac{\lambda}{2} \|x_n - u\|_{\Phi}^2 \\ &\quad \text{(H}_4\text{)} \end{aligned}$$

$$\frac{\lambda}{2} \|x_n - u\|_{\Phi}^2 \leq \underbrace{f(x_n) - f(u)}_{\substack{\downarrow \\ 0}} - \underbrace{(\nabla f(u), x_n - u)_{\Phi}}_{\substack{\downarrow \text{ (Th. 2.2.1.1) } \\ 0}} \quad \text{convergence faible}$$

$\lambda > 0 \implies x_n \xrightarrow{} u$ dans Φ fort.

II.2.2.- Algorithme stochastique

Le problème est le même qu'au § II.2.1, mais cette fois, l'observation nous fournit seulement une réalisation de la variable aléatoire $\nabla f(x) + Z(x)$ pour tout x où $Z(x)$ est une variable aléatoire à valeurs dans Φ .

Hypothèses sur le bruit :

$$\underline{H_5} : E [P(x - \rho(\nabla f(x) + Z(x)))] = P(x - \rho \nabla f(x))$$

$$\underline{H_6} : E \left[\|Z(x)\|_{\Phi}^2 \right] \leq B^2 .$$

Remarques : 1) Dans le cas sans contrainte ($K = \Phi$)

$$\underline{H_5} \iff E [Z(x)] = 0$$

2) Dans le cas où K est un sous-espace fermé de Φ (P linéaire)

$$\underline{H_5} \iff E [P Z(x)] = 0$$

$$\text{et } E [Z(x)] = 0 \implies \underline{H_5}$$

3) Il est difficile de trouver d'autres conditions générales assurant la validité de $\underline{H_5}$. Dans les exemples, la nature physique du problème et la méthode expérimentale fournissant les observations peuvent rendre $\underline{H_5}$ raisonnable même dans le cas non linéaire. S'il n'en est pas ainsi, on pourra utiliser d'autres approches que la méthode du gradient avec projection. En ce cas, nous étudierons par la suite les méthodes de pénalisation et de dualité.

Définition de l'algorithme :

$$X_0 \in L^2(\underline{\omega}, \Phi) \quad (\text{par exemple un élément fixe de } K)$$

$$X_{n+1} = P(X_n - \rho_n (\nabla f(X_n) + Z_n))$$

où Z_n est une variable aléatoire à valeurs dans Φ dont la loi conditionnelle à $X_n = x$ fixé est celle de $Z(x)$, $\rho_n > 0$,

$$\sum_n \rho_n = +\infty, \quad \sum_n \rho_n^2 < +\infty .$$

Remarque : Si on remplace $\underline{H_4}$, $\lambda > 0$ par : f strictement convexe (cf. remarque 2 § II.1.2) on doit remplacer $\sum_n \rho_n^2 < +\infty$ par $\sum_n \rho_n \varphi(\rho_n) < +\infty \iff \sum_n \delta_M(\rho_n) < +\infty$

Montrons qu'ainsi défini, $X_n \in L^2(\underline{\omega}, \Phi) \quad \forall n$

$$X_0 = x_0 \in L^2(\underline{\omega}, \Phi) \quad (X_0 \text{ non aléatoire})$$

$$\implies \nabla f(X_0) = \nabla f(x_0) \in L^2(\underline{\omega}, \Phi)$$

$$Z_0 = Z(x_0) \in L^2(\underline{\omega}, \Phi) \implies X_1 = P(X_0 - \rho_0 (\nabla f(X_0) + Z_0)) \in L^2(\underline{\omega}, \Phi) .$$

Hypothèse de récurrence :

$$X_n, X_{n-1}, \nabla f(X_{n-1}) \in L^2(\underline{\omega}, \Phi) .$$

Alors $\underline{H}_3 : \|\nabla f(X_n) - \nabla f(X_{n-1})\|_{\Phi} \leq \Lambda \|X_n - X_{n-1}\|_{\Phi}$

$$\implies \nabla f(X_n) \in L^2(\underline{\omega}, \Phi)$$

$$\|X_{n+1} - X_n\|_{\Phi}^2 \leq \rho_n^2 \|\nabla f(X_n) + Z_n\|_{\Phi}^2 \quad \text{d'après (1.5.3)}$$

$$\implies X_{n+1} \in L^2(\underline{\omega}, \Phi) .$$

Convergence

Théorème 2.2.2.1.-

$$\text{Sous } \underline{H}_1, \dots, \underline{H}_6 \quad E[f(X_n)] \xrightarrow{n \rightarrow \infty} L = \inf_{x \in K} f(x)$$

Etablissons d'abord :

Lemme 2.2.2.1.-

$$E[f(X_n)] \xrightarrow{n \rightarrow \infty} L'$$

Démonstration :

$$(2.2.1) \quad f(X_n) - f(X_{n+1}) = -(\nabla f(X_n), X_{n+1} - X_n)_{\Phi} - \frac{\nu_n}{2} \|X_{n+1} - X_n\|_{\Phi}^2$$

$$(\text{comme pour démontrer le lemme 2.2.1.1}) \quad \lambda \leq \nu_n \leq \Lambda$$

$$E\{-\rho_n(\nabla f(X_n), X_{n+1} - X_n)_{\Phi} \mid X_n\} = (X_n - \rho_n \nabla f(X_n) - X_n, E[P(X_n - \rho_n(\nabla f(X_n) + Z_n)) \mid X_n] - PX_n)_{\Phi}$$

(Prop. 1.4.1)

$$= \underbrace{(X_n - \rho_n \nabla f(X_n) - X_n, P(X_n - \rho_n \nabla f(X_n)) - PX_n)_{\Phi}}_{\substack{(H_5) \\ \geq 0}}$$

$$\implies E[-\rho_n(\nabla f(X_n), X_{n+1} - X_n)_{\Phi}] = -\rho_n(\nabla f(X_n), X_{n+1} - X_n)_{L^2} \geq 0$$

$$\|\sqrt{\nu_n}(X_{n+1} - X_n)\|_{L^2}^2 = \nu'_n \|X_{n+1} - X_n\|_{L^2}^2 \quad \lambda \leq \nu'_n \leq \Lambda$$

$$\|X_{n+1} - X_n\|_{\Phi}^2 = -\theta_n(\rho_n(\nabla f(X_n) + Z_n), P(X_n - \rho_n(\nabla f(X_n) + Z_n)) - PX_n)_{\Phi}$$

$$0 \leq \theta_n \leq 1 \quad \text{d'après (1.5.2)}$$

$$\text{et } \|X_{n+1} - X_n\|_{L^2}^2 = -\theta'_n \rho_n (\nabla f(X_n) + Z_n, X_{n+1} - X_n)_{L^2} \quad 0 \leq \theta'_n \leq 1.$$

$$\text{Donc } E[f(X_n)] - E[f(X_{n+1})] =$$

$$= \underbrace{-\rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} \left[\frac{1}{\rho_n} - \frac{v'_n \theta'_n}{2} \right]}_{v_n} - \underbrace{\frac{v'_n \theta'_n}{2} [-\rho_n (Z_n, X_{n+1} - X_n)_{L^2}]}_{w_n}$$

$$\left[\frac{1}{\rho_n} - \frac{v'_n \theta'_n}{2} \right] \geq \frac{1}{\rho_n} - \frac{\Lambda}{2} > 0 \text{ pour } n \geq N_0 \text{ car } \rho_n \xrightarrow[n \rightarrow \infty]{} 0$$

$$\implies v_n \geq 0$$

$$\begin{aligned} -\rho_n (Z_n, X_{n+1} - X_n)_{L^2} &= (-\rho_n Z_n, P(X_n - \rho_n (\nabla f(X_n) + Z_n))) - P(X_n)_{L^2} \\ &= (-\rho_n Z_n, P(X_n - \rho_n (\nabla f(X_n) + Z_n)) - P(X_n - \rho_n \nabla f(X_n)))_{L^2}. \end{aligned}$$

Car $P(X_n - \rho_n \nabla f(X_n))$ est une fonction mesurable de X_n

$$\leq \rho_n^2 \|Z_n\|_{L^2}^2 \leq \rho_n^2 B^2 \quad (\underline{H_6})$$

$$|w_n| \leq \frac{\Lambda}{2} B^2 \rho_n^2 \implies \sum_n |w_n| < +\infty$$

$$\underline{H_1} \implies \{E[f(X_n)]\} \text{ est borné inférieurement.}$$

Les hypothèses de la proposition 1.5.2.2 sont vérifiées.

$$\text{Donc } \sum_n v_n < +\infty \text{ et } E[f(X_n)] \xrightarrow[n \rightarrow \infty]{} L'.$$

Lemme 2.2.2.2.-

$$\begin{aligned} &-\rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} \\ &\text{et } -(\nabla f(X_n), X_{n+1} - X_n)_{L^2} \\ &\text{sont les termes généraux } \geq 0 \text{ de séries convergentes.} \end{aligned}$$

Démonstration :

$$\text{On vient de voir que } \sum_n v_n < +\infty$$

$$\left(\frac{1}{\rho_n} - \frac{v'_n \theta'_n}{2} \right) \geq 1 \text{ pour } n \geq N_1$$

$$\implies \sum_n -\rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} < +\infty$$

et par différence $\sum_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} < +\infty$.

Lemme 2.2.2.3.-

$\{X_n\}$ est borné dans $L^2(\underline{\omega}, \Phi)$

Démonstration :

$$\begin{aligned} \|X_{n+1} - X_n\|_{L^2}^2 &= -\theta'_n \rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} - \theta'_n \rho_n (Z_n, X_{n+1} - X_n)_{L^2} \\ &\leq -\rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} + \rho_n^2 B^2 \\ \implies \sum_n \|X_{n+1} - X_n\|_{L^2}^2 &< +\infty. \end{aligned}$$

Démonstration du théorème 2.2.2.1 :

A cause du lemme 2.2.2.1, il suffit de trouver une sous-suite $\{x_{n_k}\}$ telle que $E[f(X_{n_k})] \xrightarrow[k \rightarrow \infty]{} L$.

Soit $\alpha > L$ fixé et $\varepsilon > 0$ tel que $\alpha > \alpha - \varepsilon > L$.

Notons $\hat{S}_\alpha = \{X \in L^2(\underline{\omega}, \Phi), X(\omega) \in K \text{ p. sûrement, } E(f(X)) \leq \alpha\}$.

Hypothèse : (dont on veut démontrer l'absurdité)

$\exists N (\geq N_0)$ tel que $n \geq N \implies X_n \notin \hat{S}_\alpha$
i.e. $E[f(X_n)] > \alpha$ puisque $X_n \in L^2(\underline{\omega}, \Phi)$ et, par construction
 $X_n(\omega) \in K \quad \forall \omega$.

Soit $X(\omega) = \tilde{x} \quad \forall \omega$ où $\tilde{x} \in S_{\alpha-\varepsilon}$ (défini dans démonstration du théorème 2.2.1.1)

$f(X) - f(X_n) \geq (\nabla f(X_n), X - X_n)_\Phi \quad (f \text{ convexe})$

$-(\nabla f(X_n), X - X_n)_{L^2} \geq E[f(X_n)] - E[f(X)] \geq \alpha - (\alpha - \varepsilon)$

ou $-(\nabla f(X_n), X - X_n)_{L^2} - \varepsilon \geq 0$.

En appliquant la proposition 1.5.1.2 :

$$\begin{aligned} \|X_{n+1} - \tilde{x}\|_{L^2}^2 &\leq \|X_n - \tilde{x}\|_{L^2}^2 - 2 \rho_n (\nabla f(X_n) + Z_n, X_{n+1} - \tilde{x})_{L^2} \\ &\leq \|X_n - \tilde{x}\|_{L^2}^2 - 2 \rho_n (\nabla f(X_n), X_{n+1} - \tilde{x})_{L^2} + 2 (-\rho_n Z_n, X_{n+1} - \tilde{x})_{L^2} \\ &\quad - 2 \rho_n (\nabla f(X_n), \tilde{x} - X_n)_{L^2} - 2 \rho_n \varepsilon \end{aligned}$$

$$(-\rho_n Z_n, X_{n+1} - \tilde{x})_{L^2} = (-\rho_n Z_n, P(X_n - \rho_n (\nabla f(X_n) + Z_n)) - P(X_n - \rho_n \nabla f(X_n)))_{L^2} \leq \rho_n^2 B^2$$

$$\|X_{n+1} - \tilde{x}\|_{L^2}^2 \leq \|X_n - \tilde{x}\|_{L^2}^2 - 2 \rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} - 2 \rho_n \varepsilon + 2 \rho_n^2 B^2.$$

Itérons cette inégalité de N à $N+m$

$$0 \leq \|X_{N+m+1} - \tilde{x}\|_{L^2}^2 \leq \|X_N - \tilde{x}\|_{L^2}^2 + 2 \sum_{n=N}^{N+m} [-\rho_n (\nabla f(X_n), X_{n+1} - X_n)_{L^2} + \rho_n^2 B^2] - 2\varepsilon \sum_{n=N}^{N+m} \rho_n$$

Le 2ème terme du 2ème membre est ≥ 0 et la série converge (lemme 2.2.2.2)

$\sum_n \rho_n = +\infty \Rightarrow$ Pour m assez grand, le 2ème membre devient < 0 , d'où l'absurdité.

Par conséquent, $\forall \alpha > L$, $\forall N$, $\exists n_\alpha \geq N$ tel que $X_{n_\alpha} \in \hat{S}_\alpha$.

Il existe donc une sous-suite X_{n_k} telle que $E[f(X_{n_k})] \xrightarrow[k \rightarrow \infty]{} L$.

Théorème 2.2.2.2.-

Sous les mêmes hypothèses et si de plus $\lambda > 0$,

$\{z \in K \mid f(z) = L\}$ est réduit à un point, soit x et $X_n \xrightarrow[n \rightarrow \infty]{} x$ en moyenne quadratique.

Démonstration :

Sous ces hypothèses, l'existence et l'unicité de x sont classiques.

. Montrons d'abord la convergence faible.

$K \subset \Phi$ convexe fermé $\Rightarrow L^2(\underline{\omega}, K) \subset L^2(\underline{\omega}, \Phi)$ convexe fermé, donc faiblement fermé dans $L^2(\underline{\omega}, \Phi)$

$\{X_n\}$ borné dans $L^2(\underline{\omega}, \Phi)$ (lemme 2.2.2.3) $\Rightarrow$

$\exists$ une sous-suite $\{X_{n_k}\}$, $X_{n_k} \xrightarrow[k \rightarrow \infty]{} V$ dans $L^2(\underline{\omega}, \Phi)$ faible

$L^2(\underline{\omega}, K)$ faiblement fermé $\Rightarrow V(\omega) \in K$ p.s.

Th. 2.2.2.1 $\Rightarrow E[f(X_{n_k})] \xrightarrow[k \rightarrow \infty]{} L$.

En posant $g : U \mapsto g(U) = E[f(U)]$

$L^2(\underline{\omega}, \Phi) \longrightarrow \mathbb{R}$.

g est convexe puisque f l'est et g est Frechet-différentiable : $\nabla g(U) = \nabla f(U)$

où $\nabla f(U) : \omega \mapsto \nabla f(U(\omega))$.

Donc : $L = \lim_{k \rightarrow \infty} E[f(X_{n_k})] = \underline{\lim}_{k \rightarrow \infty} E[f(X_{n_k})] \geq E[f(U)]$.

Mais $V(\omega) \in K$ p.s. $\implies f(V(\omega)) \geq L$ p.s. $\implies E[f(V)] \geq L$.

Donc $E[f(V)] = L \implies V(\omega) = x$ p.s. .

D'après l'unicité de x , la limite faible est indépendante de la sous-suite choisie et $X_n \xrightarrow[n \rightarrow \infty]{} x$ dans $L^2(\omega, \Phi)$ faible.

. Montrons la convergence forte :

$$f(X_n) - L = f(X_n) - f(x) \geq (\nabla f(x), X_n - x)_{\Phi} + \frac{\lambda}{2} \|X_n - x\|_{\Phi}^2 .$$

$$\begin{aligned} \text{En intégrant : } \frac{\lambda}{2} \|X_n - x\|_{L^2}^2 &\leq \underbrace{E[f(X_n)] - L}_{\downarrow \text{Th. 2.2.2.1} \atop 0} - \underbrace{(\nabla f(x), X_n - x)_{L^2}}_{\downarrow \text{convergence faible} \atop 0} \end{aligned}$$

$$\lambda > 0 \implies X_n \xrightarrow[n \rightarrow \infty]{} x \text{ en moyenne quadratique.}$$

Pour obtenir un théorème de convergence presque sûre, nous allons préciser les conditions de réalisation des observations.

Nous avons supposé jusqu'ici que le bruit en un point de mesure x , ne dépend que de la valeur de x . Nous supposons maintenant que le bruit est indépendant de toutes les mesures précédentes, ce qui revient à dire que la loi de Z_n à $X_1, \dots, X_n$ fixés est identique à la loi à $X_n = x$ fixé, qui était elle-même la loi de $Z(x)$. Cette condition ne diminue pas la généralité et on a alors le :

Théorème 2.2.2.3.-

Sous les hypothèses du Théorème 2.2.2.2,

$$X_n \xrightarrow[n \rightarrow +\infty]{} x \text{ presque sûrement.}$$

Démonstration :

x réalise le minimum de f sur K . Donc $x = P(x - \rho \nabla f(x)) \quad \forall \rho > 0$.

$$\|X_{n+1} - x\|_{\Phi}^2 = \|P(X_n - \rho_n (\nabla f(X_n) + Z_n)) - P(x - \rho_n \nabla f(x))\|_{\Phi}^2 .$$

D'après la proposition 1.5.1.2,

$$\|X_{n+1} - x\|_{\Phi}^2 \leq \|X_n - x - \rho_n (\nabla f(X_n) - \nabla f(x))\|_{\Phi}^2 + 2(-\rho_n Z_n, X_{n+1} - x)_{\Phi}$$

$$\underline{H_3}, \underline{H_4} \implies \|X_{n+1} - x\|_{\Phi}^2 \leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|_{\Phi}^2 + 2(-\rho_n Z_n, X_{n+1} - x)_{\Phi}$$

$\rho_n \xrightarrow[n \rightarrow \infty]{} 0$. Donc pour n assez grand :

$$\|X_{n+1} - x\|_{\Phi}^2 \leq \|X_n - x\|_{\Phi}^2 + 2(-\rho_n Z_n, X_{n+1} - x)_{\Phi}$$

$$E \left[\|X_{n+1} - x\|_{\Phi}^2 \mid X_1, \dots, X_n \right] \leq \|X_n - x\|_{\Phi}^2 + 2 \left[E \left[(-\rho_n Z_n, X_{n+1} - x)_{\Phi} \mid X_1, \dots, X_n \right] \right]$$

$$\begin{aligned} E \left[(-\rho_n Z_n, X_{n+1} - x)_{\Phi} \mid X_1, \dots, X_n \right] &= E \left[(-\rho_n Z_n, P(X_n - \rho_n (\nabla f(X_n) + Z_n)) \right. \\ &\quad \left. - P(X_n - \rho_n \nabla f(X_n)))_{\Phi} \mid X_1, \dots, X_n \right] \\ &\leq \rho_n^2 B^2 \end{aligned}$$

$$E \left[\|X_{n+1} - x\|_{\Phi}^2 \mid X_1, \dots, X_n \right] \leq \|X_n - x\|_{\Phi}^2 + 2\rho_n^2 B^2 .$$

$$\text{Posons } U_n = -\|X_n - x\|_{\Phi}^2 + 2 B^2 \sum_{k=0}^{n-1} \rho_k^2$$

$\mathcal{a}_n$ = Tribu engendrée par $X_1, \dots, X_n$

$$\mathcal{a}_n \subset \mathcal{a}_{n+1} \quad \text{et} \quad E^{\mathcal{a}_{n+1}}(U_{n+1}) \geq U_n$$

$$E(|U_n|) \leq \|X_n - x\|_{L^2}^2 + 2 B^2 \sum_{k=0}^{n-1} \rho_k^2 \text{ borné } \left(\sum_0^{\infty} \rho_k^2 < +\infty \text{ et th. 2.2.2.2) .} \right.$$

On peut donc utiliser le théorème de convergence des sous-martingales (Doob [1]) : $\exists U$ tel que $U_n \xrightarrow[n \rightarrow \infty]{} U$ presque sûrement.

Mais $X_n \xrightarrow[n \rightarrow \infty]{} x$ en moyenne quadratique

$$\implies \exists \text{ une sous-suite } \{X_{n_\ell}\} \text{ telle que } X_{n_\ell} \xrightarrow[\ell \rightarrow \infty]{} x \text{ p.s.}$$

(démonstration identique à celle qu'on fait pour les variables aléatoires réelles en passant par la convergence en probabilité pour des variables aléatoires à valeurs dans Φ définie par :

$$Y_n \xrightarrow[n \rightarrow \infty]{} Y \text{ en proba } \iff \forall \varepsilon > 0 \quad P(\|Y_n - Y\|_{\Phi} > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0)$$

$$\|X_{n_\ell} - x\|_\Phi^2 \xrightarrow{\ell \rightarrow \infty} 0 \text{ p.s.} \implies 2 B^2 \sum_{k=0}^{n_\ell} \rho_k^2 \xrightarrow{\ell \rightarrow \infty} U \text{ p.s.}$$

$$\implies (\text{prop. 1.5.2.1}) \quad 2 B^2 \sum_{k=0}^n \rho_k^2 \xrightarrow{n \rightarrow \infty} U \text{ p.s.}$$

$$\text{c'est-à-dire } U = 2 B^2 \sum_{k=0}^{\infty} \rho_k^2 \text{ p.s.}$$

$$\implies \|X_n - x\|_\Phi^2 \xrightarrow{n \rightarrow \infty} 0 \text{ p.s. i.e. } X_n \xrightarrow{n \rightarrow \infty} x \text{ p.s.}$$

II.3.- Algorithme de Gradient stochastique

II.3.1.- Définition de l'algorithme

On se place dans le cas d'une observation stochastique définie comme au § II.1.1.

La différence avec l'algorithme du § II.2 tient dans les propriétés du pas d'itération ρ_n .

$$\left| \begin{array}{l} X_0 \in L^2(\omega, \Phi) \quad (\text{par exemple un élément fixe de } K) \\ X_{n+1} = P(X_n - \rho_n (\nabla f(X_n) + Z_n)) \\ Z_n \text{ est défini comme au § II.2.2.} \\ \delta \leq \rho_n \leq \frac{2}{\Lambda} - \delta \quad \text{avec } \delta \geq \frac{2}{\Lambda} \frac{B}{1+B} > 0 \end{array} \right.$$

(λ et Λ sont définies dans $\underline{H_3}$ et $\underline{H_4}$ § II.1.2)

B est définie ci-dessous.

Au § II.2.2, un facteur déterminant de la convergence de l'algorithme était que $\sum_n \rho_n^2 \|Z_n\|_{L^2}^2 < +\infty$.

Sous l'hypothèse $\underline{H_6}$ faite alors, ceci n'est à présent plus vérifié car $\rho_n \geq \delta > 0$.

On devra donc remplacer l'hypothèse $\underline{H_6}$ par une autre hypothèse exprimant l'idée intuitive qu'en moyenne quadratique, le bruit est borné en valeur relative par rapport au gradient.

$\underline{H_5}$ ne sera alors plus nécessaire, il suffira de

$$\underline{H'_5} \quad E(Z(x)) = 0$$

$\underline{H_6}$ est remplacé par :

$$\underline{H'_6} \quad E \left[(Z(x), P_x - P(x - \rho(\nabla f(x) + Z(x))))_{\Phi} \right] \leq \\ B \quad E \left[(\nabla f(x), P_x - P(x - \rho(\nabla f(x) + Z(x))))_{\Phi} \right]$$

$$\underline{H_7} \quad 0 \leq B \leq 1$$

ou selon le cas

$$\underline{H'_7} \quad 0 \leq B \leq \frac{1}{2\Lambda - \lambda} \leq 1 \quad .$$

Remarques : - L'hypothèse $\underline{H'_6}$ représente bien une hypothèse sur le rapport signal bruit. Si le problème est sans contrainte ($K = \Phi$), $\underline{H'_6}$ se réduit à $E \left[\|Z(x)\|_{\Phi}^2 \right] \leq B \|\nabla f(x)\|_{\Phi}^2$.

- Sous l'hypothèse $\underline{H_7}$ (ou $\underline{H'_7}$) le choix $\delta \geq \frac{2}{\Lambda} \frac{B}{1+B}$ est possible

$$\text{et on a } \frac{2}{\Lambda} \frac{B}{1+B} \leq \delta \leq \rho_n \leq \frac{2}{\Lambda} - \delta \leq \frac{2}{\Lambda(1+B)}$$

- Comme au § II.2.2 $X_n \in L^2(\underline{\omega}, \Phi) \quad \forall n$.

II.3.2.- Convergence de l'algorithme

Théorème 2.3.2.1.-

$$\text{Sous } \underline{H_1}, \underline{H_2}, \underline{H_3}, \underline{H_4}, \underline{H'_5}, \underline{H'_6}, \underline{H'_7}, \\ E[f(X_n)] \xrightarrow{n \rightarrow \infty} L = \inf_{x \in K} f(x)$$

Etablissons d'abord le

Lemme 2.3.2.1.-

$$E[f(X_n)] \xrightarrow{n \rightarrow \infty} L'$$

Démonstration :

La formule (2.2.1) est valable et donne :

$$(2.3.1) \quad f(X_n) - f(X_{n+1}) \geq -(\nabla f(X_n), X_{n+1} - X_n) - \frac{\Lambda}{2} \|X_{n+1} - X_n\|_{\Phi}^2$$

$$- \rho_n (\nabla f(X_n), X_{n+1} - X_n)_\Phi = (X_n - \rho_n (\nabla f(X_n) + Z_n) - X_n, P(X_n - \rho_n (\nabla f(X_n) + Z_n) - PX_n)_\Phi \\ + (\rho_n Z_n, X_{n+1} - X_n)_\Phi .$$

$$(2.3.2) \quad - (\nabla f(X_n), X_{n+1} - X_n)_\Phi = \frac{\mu^2(\rho_n)}{\rho_n} \|X_{n+1} - X_n\|_\Phi^2 + (Z_n, X_{n+1} - X_n)_\Phi$$

où $\mu^2(\rho_n) \geq 1$ d'après (1.5.2)

$\mu^2(\rho_n)$ étant ainsi défini, (2.3.1) implique

$$(2.3.3) \quad f(X_n) - f(X_{n+1}) \geq -(\nabla f(X_n), X_{n+1} - X_n)_\Phi - \frac{\Lambda}{2} \mu^2(\rho_n) \|X_{n+1} - X_n\|_\Phi^2 .$$

En utilisant $\underline{H}_6$ et H_3 qui implique que $\nabla f(\cdot)$ est continue, donc mesurable ainsi que la proposition (1.4.2.3) :

$$E \left[(Z_n, X_{n+1} - X_n)_\Phi \right] = E \left[E \left[(Z_n, P(X_n - \rho_n (\nabla f(X_n) + Z_n)) - PX_n)_\Phi \mid X_n \right] \right] \\ \geq B E \left[E \left[(\nabla f(X_n), P(X_n - \rho_n (\nabla f(X_n) + Z_n)) - PX_n)_\Phi \mid X_n \right] \right]$$

$$(2.3.4) \quad E \left[(Z_n, X_{n+1} - X_n)_\Phi \right] \geq B E \left[(\nabla f(X_n), X_{n+1} - X_n)_\Phi \right] .$$

De (2.3.2) et (2.3.4) , on tire

$$- (\nabla f(X_n), X_{n+1} - X_n)_{L^2} \geq \frac{1}{\rho_n} \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 + B(\nabla f(X_n), X_{n+1} - X_n)_{L^2}$$

$$\text{ou } - (1+B) (\nabla f(X_n), X_{n+1} - X_n)_{L^2} \geq \frac{1}{\rho_n} \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 .$$

En portant dans (2.3.3) où on a pris l'espérance mathématique :

$$(2.3.5) \quad E [f(X_n)] - E [f(X_{n+1})] \geq \left[\frac{1}{\rho_n(1+B)} - \frac{\Lambda}{2} \right] \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2$$

$$\rho_n \leq \frac{2}{\Lambda} - \delta \leq \frac{1}{\Lambda(1+B)} \implies \frac{1}{\rho_n(1+B)} - \frac{\Lambda}{2} \geq 0 .$$

Donc la suite $E [f(X_n)]$ est décroissante, bornée inférieurement ($\underline{H}_1$) .

Lemme 2.3.2.2.-

$$\|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 \text{ et } - (\nabla f(X_n), X_{n+1} - X_n)_{L^2} \text{ sont}$$

les termes généraux ≥ 0 de séries convergentes

Démonstration :

La série de terme général $E[f(X_n)] - E[f(X_{n+1})]$ converge.

De plus $-(\nabla f(X_n), X_{n+1} - X_n)_{L^2} \geq \frac{1}{\rho_n(1+B)} \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 \geq 0$

et $-(\nabla f(X_n), X_{n+1} - X_n) \leq E[f(X_n)] - E[f(X_{n+1})] + \frac{A}{2} \|X_{n+1} - X_n\|_{L^2}^2$.

Lemme 2.3.2.3.-

$$\{X_n\} \text{ est borné dans } L^2(\underline{\omega}, \Phi)$$

Corollaire immédiat du lemme 2.3.2.2 puisque $\sum_n \|X_{n+1} - X_n\|_{L^2}^2 < +\infty$.

Démonstration du théorème 2.3.2.1 :

Il suffit de montrer que $L' = L$.

Supposons donc $L' > L$. Soit $z \in K$ tel que $f(z) < L'$.

$$0 > -\varepsilon = f(z) - L' \geq f(z) - E[f(X_n)] \geq (\nabla f(X_n), z - X_n)_{L^2} \quad (f \text{ convexe})$$

$$0 > -\varepsilon \delta \geq (\rho_n \nabla f(X_n), z - X_n)_{L^2} = \rho_n (\nabla f(X_n) + Z_n, z - X_n)_{L^2} \quad (\underline{H}_5' \text{ et prop. (1.4.2.3)})$$

$$= \rho_n (\nabla f(X_n) + Z_n, z - X_{n+1} + X_{n+1} - X_n)_{L^2}.$$

En utilisant (2.3.2)

$$0 > -\varepsilon \delta \geq - \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 + \rho_n (\nabla f(X_n) + Z_n, z - X_{n+1})_{L^2}$$

$$(2.3.6) \quad 0 > -\varepsilon \delta \geq - \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 + (X_n - X_{n+1}, z - X_{n+1})_{L^2} + (\rho_n (\nabla f(X_n) + Z_n) + X_{n+1} - X_n, z - X_{n+1})_{L^2}.$$

En posant $U = X_n - \rho_n (\nabla f(X_{n+1}) + Z_n)$, $X_{n+1} = PU$

et le 2ème terme du 2ème membre de (2.3.6) s'écrit

$$(PU - U, z - PU)_{L^2} \geq 0 \text{ d'après (1.5.1).}$$

$$\text{Donc (2.3.7) } 0 > -\varepsilon \delta \geq - \|\mu(\rho_n)(X_{n+1} - X_n)\|_{L^2}^2 + (X_n - X_{n+1}, z - X_{n+1})_{L^2}.$$

D'après le lemme 2.3.2.3, les X_n sont bornés dans L^2 .

En passant à la limite dans (2.3.7) et en utilisant le lemme 2.3.2.2, $0 > -\varepsilon\delta > 0$ absurde.

D'où $L' = L$.

Théorème 2.3.2.2.-

Sous les mêmes hypothèses sauf H_7' remplaçant H_7 et si de plus $\lambda > 0$, $\delta \leq \rho_n \leq \frac{2\lambda}{\Lambda^2} - \delta$

- $\{z \in K \mid f(z) = L\}$ est réduit à un point, soit x

- et $X_n \xrightarrow[n \rightarrow \infty]{} x$ en moyenne quadratique.

Démonstration :

Sous ces hypothèses, l'existence et l'unicité de x sont classiques.

H_7' et $\lambda > 0$ impliquent $\delta \leq \frac{\lambda}{\Lambda^2}$ et on a

$$\frac{2}{\Lambda} \frac{B}{1+B} \leq \delta \leq \rho_n \leq \frac{2\lambda}{\Lambda^2} - \delta \leq \frac{2}{\Lambda} - \delta \leq \frac{2}{\Lambda(1+B)}$$

x vérifie $x = P(x - \rho \nabla f(x))$

et d'après la proposition 1.5.1.2

$$\|X_{n+1} - x\|_{\Phi}^2 \leq \|X_n - x - \rho_n (\nabla f(X_n) - \nabla f(x))\|_{\Phi}^2 - 2(\rho_n Z_n, X_{n+1} - x)_{\Phi}$$

$$\underline{H_3}, \underline{H_4} \implies \|X_{n+1} - x\|_{\Phi}^2 \leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|_{\Phi}^2 - 2(\rho_n Z_n, X_{n+1} - x)_{\Phi}$$

$$0 < \max_{\delta \leq \rho \leq \frac{2\lambda}{\Lambda^2} - \delta} (1 - 2\rho\lambda + \rho^2 \Lambda^2) = 1 - 2\lambda\delta + \delta^2 \Lambda^2 = 1 - \alpha < 1$$

$$\text{car } \delta \leq \frac{\lambda}{\Lambda^2} < \frac{2\lambda}{\Lambda^2} \text{ puisque } \lambda > 0.$$

Donc :

$$(2.3.8) \quad E \left[\|X_{n+1} - x\|_{\Phi}^2 \mid X_n \right] \leq (1 - \alpha) \|X_n - x\|_{\Phi}^2 + 2 E \left[(\rho_n Z_n, x - X_{n+1}) \mid X_n \right]$$

$$(\rho_n Z_n, x - X_{n+1})_{\Phi} = (\rho_n Z_n, x - X_n)_{\Phi} + (\rho_n Z_n, X_n - X_{n+1})_{\Phi}$$

$$\begin{aligned} E \left[(\rho_n Z_n, x - X_{n+1})_{\Phi} \mid X_n \right] &= 0 + E \left[(\rho_n Z_n, X_n - X_{n+1})_{\Phi} \mid X_n \right] \\ &\leq \rho_n B E \left[(\nabla f(X_n), X_n - X_{n+1})_{\Phi} \mid X_n \right] \end{aligned}$$

$$\text{et } 2(\rho_n Z_n, x - X_{n+1})_{\Phi} \leq 2\rho_n B (\nabla f(X_n), X_n - X_{n+1})_{\Phi} = \beta_n.$$

En prenant l'espérance mathématique dans (2.3.8)

$$(2.3.9) \quad \|X_{n+1} - x\|_{L^2}^2 \leq (1-\alpha) \|X_n - x\|_{L^2}^2 + \beta_n$$

où $\beta_n \geq 0$ est le terme général d'une série convergente (lemme 2.3.2.2) .

En itérant (2.3.9) :

$$\|X_{n+1} - x\|_{L^2}^2 \leq (1-\alpha)^n \|X_1 - x\|_{L^2}^2 + \beta_n + \beta_{n-1} (1-\alpha) + \dots + \beta_1 (1-\alpha)^{n-1} .$$

Le 2ème membre est le terme général d'une série convergente

$$\implies \|X_n - x\|_{L^2}^2 \xrightarrow{n \rightarrow \infty} 0 .$$

Théorème 2.3.2.3.-

Sous les hypothèses du théorème 2.3.2.2,

$$X_n \xrightarrow{n \rightarrow \infty} x \text{ presque sûrement.}$$

Démonstration : (à l'aide de la théorie des martingales)

$$\rho_n (\nabla f(X_n) + Z_n, X_n - X_{n+1})_{\Phi} = \mu^2(\rho_n) \|X_{n+1} - X_n\|_{\Phi}^2 \geq 0 \quad \mu^2(\rho_n) \geq 1 .$$

Donc

$$\rho_n (1+B) E \left[(\nabla f(X_n), X_n - X_{n+1})_{\Phi} | X_1 \dots X_n \right] \geq \rho_n E \left[(\nabla f(X_n) + Z_n, X_n - X_{n+1})_{\Phi} | X_1 \dots X_n \right] \geq 0 \quad (1)$$

$$(\rho_n Z_n, x - X_{n+1})_{\Phi} = (\rho_n Z_n, x - X_n)_{\Phi} + \rho_n (\nabla f(X_n) + Z_n, X_n - X_{n+1})_{\Phi} - \rho_n (\nabla f(X_n), X_n - X_{n+1})_{\Phi}$$

$$= (\rho_n Z_n, x - X_n)_{\Phi} + \mu^2(\rho_n) \|X_n - X_{n+1}\|_{\Phi}^2 - \rho_n (\nabla f(X_n), X_n - X_{n+1})_{\Phi}$$

$$\begin{aligned} E \left[(\rho_n Z_n, x - X_{n+1})_{\Phi} | X_1 \dots X_n \right] &= 0 + E \left[\mu^2(\rho_n) \|X_n - X_{n+1}\|_{\Phi}^2 | X_1 \dots X_n \right] \\ &\quad - E \left[(\rho_n \nabla f(X_n), X_n - X_{n+1})_{\Phi} | X_1 \dots X_n \right] \\ &\leq E \left[\mu^2(\rho_n) \|X_n - X_{n+1}\|_{\Phi}^2 | X_1 \dots X_n \right] . \end{aligned}$$

(1) La remarque faite avant le théorème 2.2.2.3 sur la définition de Z_n est encore valable.

(2.3.8) pouvait aussi bien s'écrire :

$$\begin{aligned} E \left[\|X_{n+1} - x\|_{\Phi}^2 \mid X_1 \dots X_n \right] &\leq (1-\alpha) \|X_n - x\|_{\Phi}^2 + 2 E \left[(\rho_n Z_n, x - X_{n+1})_{\Phi} \mid X_1 \dots X_n \right] \\ &\leq \|X_n - x\|_{\Phi}^2 + 2 E \left[\mu^2(\rho_n) \|X_n - X_{n+1}\|_{\Phi}^2 \mid X_1 \dots X_n \right] \end{aligned}$$

$$\begin{aligned} \text{ou } E \left\{ \left[\|X_{n+1} - x\|_{\Phi}^2 - 2 \sum_{k=1}^n \mu^2(\rho_k) \|X_k - X_{k+1}\|_{\Phi}^2 \right] \mid X_1 \dots X_n \right\} \\ \leq \|X_n - x\|_{\Phi}^2 - 2 \sum_{k=1}^{n-1} \mu^2(\rho_k) \|X_k - X_{k+1}\|_{\Phi}^2 = - U_n \end{aligned}$$

c'est-à-dire avec les notations du théorème 2.2.2.3

$$E^{Q_{n(U_{n+1})}} \geq U_n$$

$$E(|U_n|) \leq \|X_n - x\|_{L^2}^2 + 2 \sum_{k=1}^{n-1} \|\mu(\rho_k) (X_k - X_{k+1})\|_{L^2}^2$$

borné (théorème 2.3.2.2 et lemme 2.3.2.2).

D'où à l'aide du théorème de convergence des sous-martingales (Doob [1])

$\exists U$ tel que $U_n \xrightarrow[n \rightarrow \infty]{} U$ presque sûrement.

Mais $X_{n_l} \xrightarrow[l \rightarrow \infty]{} x$ dans $L^2(\omega, \Phi) \implies \exists$ une sous-suite $\{X_{n_l}\} : X_{n_l} \xrightarrow[l \rightarrow \infty]{} x$ p.s.

Donc $\|X_{n_l} - x\|_{\Phi}^2 \xrightarrow[l \rightarrow \infty]{} 0$ p.s.

et $-2 \sum_{k=1}^{n_l-1} \mu^2(\rho_k) \|X_k - X_{k+1}\|_{\Phi}^2 \xrightarrow[l \rightarrow \infty]{} U$ p.s.

donc aussi (prop. 1.5.2.1) : $-2 \sum_{k=1}^{n-1} \mu^2(\rho_k) \|X_k - X_{k+1}\|_{\Phi}^2 \xrightarrow[n \rightarrow \infty]{} U$ p.s.

et par suite $X_n \xrightarrow[n \rightarrow \infty]{} x$ presque sûrement.

II.3.3.- Affaiblissement des hypothèses sur le bruit

Par rapport au théorème de gradient déterministe, les théorèmes (2.3.2.1) (2.3.2.2) (2.3.2.3) représentent une meilleure approche de la réalité. On ne suppose plus que les mesures sont faites sans erreur.

Pourtant, l'hypothèse H'_6 a peu de chances d'être vérifiée dans la pratique. Elle implique en particulier qu'au point x , abscisse du minimum, le gradient est mesuré sans erreur, ce qui n'est pas réaliste.

En revanche, l'hypothèse $\underline{H_6}$ faite dans le cadre du § II.2.2 est elle, raisonnable pour les applications.

Mais on ne peut plus espérer alors pour l'algorithme de gradient, un résultat de convergence aussi fort que précédemment. On peut démontrer le

Théorème 2.3.2.4.-

Si $\underline{H_1}, \underline{H_2}, \underline{H_3}, \underline{H_4}, \underline{H'_5}, \underline{H_6}$ sont vérifiées, si $\lambda > 0$ et si pour $n \geq N$, $\rho_n \leq \frac{\lambda}{\Lambda^2}$, alors, $\limsup_n E(\ X_n - x\ _{\Phi}^2) \leq \frac{B^2}{\Lambda^2} .$
--

Démonstration :

Les hypothèses faites assurent l'existence et l'unicité de l'abscisse x du minimum.

Appliquons la proposition 1.5.1.3 avec $a = X_n - \rho_n \nabla f(X_n)$

$$u = - \rho_n Z_n$$

$$b = x - \rho_n \nabla f(x)$$

$$\text{Alors } \|a-b\|_{\Phi}^2 \leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|_{\Phi}^2$$

$$\|X_{n+1} - x\|_{\Phi}^2 \leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|_{\Phi}^2 - 2(\rho_n Z_n, X_n - x - \rho_n (\nabla f(X_n) - \nabla f(x)))_{\Phi} + \rho_n^2 \|Z_n\|_{\Phi}^2 .$$

$$\text{D'où } E[\|X_{n+1} - x\|_{\Phi}^2 | X_n] \leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|_{\Phi}^2 + \rho_n^2 E(\|Z_n\|_{\Phi}^2 | X_n)$$

soit avec $\underline{H_6}$ puis en prenant l'espérance mathématique :

$$(2.3.10) \quad \|X_{n+1} - x\|_{L^2}^2 \leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|_{L^2}^2 + \rho_n^2 B^2 .$$

$$. \text{ Soit } n \geq N . \text{ Alors } \rho_n \leq \frac{\lambda}{\Lambda^2} \leq \frac{1}{\Lambda}$$

Soit $\varepsilon > 0$. Choisissons $n(\varepsilon) > n$ tel que

$$\varepsilon \lambda \sum_{k=n}^{n(\varepsilon)-1} \rho_k \geq \|X_n - x\|_{L^2}^2 - \frac{B^2}{\Lambda^2} . \text{ C'est possible car } \rho_k \geq \delta > 0 .$$

. S'il existe $k \geq N$ tel que

$$(2.3.11) \quad \|X_k - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} + \varepsilon ,$$

montrons que cette inégalité est vraie pour tout indice supérieur.

Reportons (2.3.11) dans (2.3.10) avec $n = k$:

$$\begin{aligned} \|X_{k+1} - x\|_{L^2}^2 &\leq (1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2) \left(\frac{B^2}{\Lambda^2} + \varepsilon \right) + \rho_k^2 B^2 \\ \|X_{k+1} - x\|_{L^2}^2 &\leq \frac{B^2}{\Lambda^2} (1 - 2\rho_k (\lambda - \rho_k \Lambda^2)) + \varepsilon (1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2) . \end{aligned}$$

Puisque $\delta \leq \rho_k \leq \frac{\lambda}{\Lambda^2}$, $1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2 < 1 - 2\rho_k (\lambda - \rho_k \Lambda^2) \leq 1$

et donc $\|X_{k+1} - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} + \varepsilon$.

. Si $\|X_k - x\|_{L^2}^2 > \frac{B^2}{\Lambda^2} + \varepsilon \quad \forall k : n \leq k \leq n(\varepsilon) - 1$, montrons que

(2.3.11) est vérifié pour $k = n(\varepsilon)$

$$(2.3.10) \text{ s'écrit } \|X_{k+1} - x\|_{L^2}^2 \leq \|X_k - x\|_{L^2}^2 - \rho_k (2\lambda - \rho_k \Lambda^2) \|X_k - x\|_{L^2}^2 + \rho_k^2 B^2$$

$$\rho_k \leq \frac{\lambda}{\Lambda^2} \implies 2\lambda - \rho_k \Lambda^2 \geq \lambda > 0 \text{ et}$$

$$\|X_{k+1} - x\|_{L^2}^2 \leq \|X_k - x\|_{L^2}^2 - \rho_k (2\lambda - \rho_k \Lambda^2) \left(\frac{B^2}{\Lambda^2} + \varepsilon \right) + \rho_k^2 B^2$$

$$\|X_{k+1} - x\|_{L^2}^2 \leq \|X_k - x\|_{L^2}^2 + \frac{2B^2}{\Lambda^2} \underbrace{(-\lambda \rho_k + \rho_k^2 \Lambda^2)}_{\leq 0} - \varepsilon \rho_k \lambda \leq \|X_k - x\|_{L^2}^2 - \varepsilon \rho_k \lambda .$$

En itérant cette inégalité entre n et $n(\varepsilon) - 1$

$$\|X_{n(\varepsilon)} - x\|_{L^2}^2 \leq \|X_n - x\|_{L^2}^2 - \varepsilon \lambda \sum_{k=n}^{n(\varepsilon)-1} \rho_k \leq \|X_n - x\|_{L^2}^2 - \|X_n - x\|_{L^2}^2 + \frac{B^2}{\Lambda^2} = \frac{B^2}{\Lambda^2} .$$

$$\text{Donc } \limsup_n \|X_n - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} .$$

Remarques : 1.- On peut trouver dans : J.P. Comer [1], un résultat du même type pour le processus de Robbins-Mourot à une dimension.

2.- Le théorème (2.3.2.4) donne une signification rigoureuse à l'opinion couramment admise selon laquelle, "la méthode du gradient marche encore si le bruit n'est pas trop grand".

Quoique souvent vérifiée, l'hypothèse H_6 est parfois mal adaptée à certains problèmes effectivement rencontrés dans la pratique. Soit par exemple, un problème

de stock où x est en excédent de stock de denrée périssable. Il est raisonnable de penser que l'espérance mathématique du coût de l'excédent de stock est fonction croissante de l'excédent x et non bornée. On développera un exemple de ce type au chapitre IV.

Pour traiter ce type de problème, démontrons le

Corollaire 2.3.2.5.-

Sous les mêmes hypothèses qu'au théorème 2.3.2.4
sauf à remplacer H_6 par :

$$\begin{aligned} E(\|Z(x)\|_{\Phi}^2) &\leq d_1 + d_2 \|x\|_{\Phi}^2 \\ \text{et } \rho_n &\leq \frac{\lambda}{\Lambda^2 + 2d_2}, \\ \text{alors } \limsup_n E(\|X_n - x\|_{\Phi}^2) &\leq \frac{d_1 + 2d_2 \|x\|_{\Phi}^2}{\Lambda^2 + 2d_2} \end{aligned}$$

Démonstration :

$$\rho_n^2 \|Z_n\|_{L^2}^2 \leq \rho_n^2 (d_1 + d_2 \|X_n\|_{L^2}^2)$$

$$\|X_n\|_{\Phi}^2 = \|X_n - x + x\|_{\Phi}^2 \leq 2 [\|X_n - x\|_{\Phi}^2 + \|x\|_{\Phi}^2].$$

$$\text{Posons } B^2 = d_1 + 2d_2 \|x\|_{\Phi}^2$$

$$\rho_n^2 \|Z_n\|_{L^2}^2 \leq 2 \rho_n^2 d_2 \|X_n - x\|_{L^2}^2 + \rho_n^2 B^2.$$

En opérant comme pour trouver (2.3.10), on en déduit :

$$\|X_{n+1} - x\|_{L^2}^2 \leq \left[1 - 2\rho_n \lambda + \rho_n^2 (\Lambda^2 + 2d_2) \right] \|X_n - x\|_{L^2}^2 + \rho_n^2 B^2.$$

On peut alors poursuivre la démonstration comme pour le théorème 2.3.2.4 en remplaçant Λ^2 par $\Lambda^2 + 2d_2$.

On peut énoncer le théorème 2.3.2.4 sous la forme :

Corollaire 2.3.2.6.-

Sous les hypothèses du théorème 2.3.2.4, $\forall \varepsilon > 0$, il existe $B(\varepsilon)$ tel que, si H_6 est vérifiée pour $B = B(\varepsilon)$, alors $\limsup_n (\|X_n - x\|_{\Phi}^2) < \varepsilon$.

ce qui exprime qu'il existe une borne supérieure de la variance du bruit permettant d'approcher la solution avec une précision donnée.

II.4.- Comparaison avec les résultats antérieurs

A notre connaissance, les résultats publiés jusqu'ici par d'autres auteurs, apportent au problème traité, les restrictions suivantes :

- Il s'agit d'un problème de minimisation sans contrainte.
- Φ est de dimension finie (Φ isomorphe à $\mathbb{R}^P$) .

Les problèmes sont pris sous l'angle de l'approximation stochastique avec un algorithme de Robbins-Monro [1] .

Le problème de l'approximation stochastique :

Il s'agit de résoudre l'équation

$g(x) = \alpha \quad x \in \Phi \quad , \quad g : \Phi \mapsto \Phi \quad , \quad \alpha \in \Phi$ connaissant la variable aléatoire à valeurs dans Φ : $Y(x) = g(x) + Z(x)$, $E(Z(x)) = 0$.

Nous nous intéresserons particulièrement au cas $g(x) = \nabla f(x)$, $\alpha = 0$.

L'algorithme de Robbins-Monro :

Il est défini par

(2.4.1) $X_{n+1} = X_n - \rho_n (g(X_n) + Z_n - \alpha)$ où Z_n est défini comme au § II.2.2 ou bien si $g(x) = \nabla f(x)$, $\alpha = 0$, $f(x)$ à valeurs réelles,

(2.4.2) $X_{n+1} = X_n - \rho_n (\nabla f(X_n) + Z_n)$

ρ_n est choisi comme au § 2.2 : $\rho_n > 0$, $\sum_n \rho_n = +\infty$, $\sum_n \rho_n^2 < +\infty$.

II.4.1.- Convergence obtenue par la théorie des martingales

Nous énonçons ci-dessous un théorème dû à Gladyshev [1] auquel nous renvoyons pour la démonstration.

Théorème 2.4.1.1.-

Soit $\theta \in \Phi$ solution de $g(x) = \alpha$

Si - $\inf_{\varepsilon < \|x - \theta\|_{\Phi} < \frac{1}{\varepsilon}} (x - \theta, g(x) - \alpha)_{\Phi} > 0 \quad \forall \varepsilon > 0$

- $\exists d > 0$ tel que pour tout $x \in \Phi$

$$E \left[\|Y(x) - \alpha\|_{\Phi}^2 \right] \leq d(1 + \|x - \theta\|_{\Phi}^2)$$

Alors X_n (défini par (2.4.1)) $\xrightarrow[n \rightarrow \infty]{} \theta$ presque sûrement et en moyenne quadratique.

En appliquant ce théorème à la résolution de $\nabla f(x) = 0$, obtient le

Corollaire 2.4.1.1.-

Soit θ le minimum de $f(x)$ sur Φ .

Si $\underline{K_1}$: f est différentiable sur Φ

$\underline{K_2}$: $\inf_{\varepsilon < \|x - \theta\|_{\Phi} < \frac{1}{\varepsilon}} (x - \theta, \nabla f(x))_{\Phi} > 0 \quad \forall \varepsilon > 0$

$\underline{K_3}$: $\exists d > 0$ tel que pour tout $x \in \Phi$

$$E \left[\|\nabla f(x) + Z(x)\|_{\Phi}^2 \right] \leq d(1 + \|x - \theta\|_{\Phi}^2)$$

Alors X_n (défini par 2.4.2) $\xrightarrow[n \rightarrow \infty]{} \theta$ presque sûrement et en moyenne quadratique.

II.4.2.- Convergence obtenue par le lemme de Borel-Cantelli

Les énoncés usuels du lemme de Borel-Cantelli sont insuffisants.

On trouve dans : Dubbins-Freedman [1], une forme plus élaborée qu'on utilise ici. On résout l'équation $\nabla f(x) = 0$, $x \in \Phi$.

Théorème 2.4.2.1.-

Si $\underline{L_1}$: f est bornée inférieurement sur Φ

$\underline{L_2}$: f admet des dérivées bornées jusqu'au
2ème ordre sur Φ

$\underline{L_3}$: θ est un minimum local de f

$\underline{L_4}$: $\forall x \in \Phi$, $x \neq \theta$, $\nabla f(x) \neq 0$

$\underline{L_5}$: $E(\|Z(x)\|^2) \leq K_0 \quad \forall x \in \Phi$

soit X_n défini en (2.4.2)

Alors presque sûrement $X_n \xrightarrow[n \rightarrow \infty]{} \theta$ ou

$\|X_n\|_{\Phi} \rightarrow +\infty$.

Si de plus :

$\underline{L_6}$: $\exists T > 0$ tel que

$\|x - \theta\|_{\Phi} > T \implies (x - \theta, \nabla f(x))_{\Phi} \geq 0$

alors $X_n \xrightarrow[n \rightarrow \infty]{} \theta$ presque sûrement

Notre démonstration du théorème 2.4.2.1 est analogue à celle du théorème 4.4.1. Nous ne la donnons donc pas ici.

II.4.3.- Comparaison entre les différents théorèmes

Cette comparaison peut être résumée dans le tableau suivant :

	Algorithme de gradient stochastique Th. (2.3.2.2) Th. (2.3.2.3)	Algorithme séries divergentes Th. (2.2.2.2) Th. (2.2.2.3)	Corollaire (2.4.1.1)	Théorème (2.4.2.1)
domaine de validité	Problème avec contraintes Φ éventuellement de dim ∞	$\longrightarrow$ idem	Problème sans contrainte Φ de dimension finie	$\longrightarrow$ idem
pas d'itération	$0 < \delta \leq \rho_n \leq \frac{2\lambda}{\Lambda^2} - \delta$	$\rho_n > 0, \sum_n \rho_n = +\infty, \sum_n \rho_n^2 < +\infty$	$\longrightarrow$ idem	$\longrightarrow$ idem
comparaison des hypothèses	H_1 H_2 $H_4 + \lambda > 0$ H'_5 $H_3 + H'_6$ (Th. (2.3.2.4) $H_3 + H'_6$)	H_1 H_2 $H_4 + \lambda > 0$ H_5 $H_3 + H_6$	comparaison	
			K_1 K_2 K_3	L_1 L_2 L_3 L_4 L_5 L_6
vitesse de convergence	rapide	lente	lente	lente
Type de convergence	presque sûre et moyenne quadratique	$\longrightarrow$ idem	$\longrightarrow$ idem	presque sûre

Ce tableau appelle les conclusions suivantes :

1 - Les théorèmes démontrés aux § 2.2 et § 2.3 ont un champ d'application plus vaste que la plupart des résultats connus jusqu'ici : Problèmes avec contraintes dans un espace de dimension infinie. Dans cette dernière direction, voir aussi J.H. Venter [1] et L. Schmetterer [1], [2] à propos des algorithmes de Robbins-Monro et de Dvoretzky.

2 - Il est bien connu que des méthodes du type steepest descent convergent plus vite que des méthodes de série divergente.

Sous des hypothèses raisonnables dans les applications, on a seulement un théorème limite (Th. 2.3.2.4) en fait suffisant numériquement.

Sous ces mêmes hypothèses, la méthode de série divergente converge encore. Naturellement, ce qu'on gagne sur le plan des hypothèses, se trouve compensé par une perte de vitesse de convergence. Si on ne désire pas se contenter des résultats de convergence du Th. 2.3.2.4, on peut adopter la démarche suivante : utiliser d'abord l'algorithme de Gradient qui nous conduira rapidement au voisinage de la solution (Th. 2.3.2.4) puis utiliser l'algorithme de série divergente qui permettra, sous les mêmes hypothèses d'affiner le résultat (Th. 2.2.2.2) avec une vitesse de convergence plus faible. On étudiera numériquement ce procédé au chapitre VI.

II.5.- Applications

Nous allons reprendre les exemples présentés au chapitre I et montrer comment on peut utiliser dans ces problèmes les résultats obtenus dans le présent chapitre.

II.5.1.- Problème de contrôle dans un système elliptique

On reprend le problème et les notations définis au § I.1.1. On sait (J.L. Lions [1]) que $\forall J(v) = 2(B^*p(v) + \Lambda_{\mathcal{U}} N v)$ où $p(v)$ (état adjoint) est solution du système adjoint :

$$A^*p(v) = \mathcal{E}^* \Lambda_{\mathcal{Y}} (\mathcal{E}y(v) - z_d) .$$

Soit $v \in \mathcal{U}$. L'observation nous fournit (une réalisation de)

$$\mathcal{E}y(v) + \eta \quad \text{où } \eta \text{ est une variable aléatoire à valeurs dans } \mathcal{U} .$$

Soit $p_1(v)$ solution de (2.5.1) $A^* p_1(v) = \mathcal{E}^* \Lambda_{\mathcal{H}} (\mathcal{E}_y(v) + \eta - z_d)$.

Alors :

$$(2.5.2) \quad 2(B^* p_1(v) + \Lambda_{\mathcal{U}} N v) = \nabla J(v) + 2 B^* (p_1(v) - p(v)) \\ = \nabla J(v) + \xi$$

où ξ est un bruit dont les propriétés stochastiques se désuisent de celles de η . Par exemple, grâce à la linéarité du système, si $E(\eta) = 0$, alors $E(\xi) = 0$.

L'observation permet donc d'obtenir, moyennant la résolution d'un seul système aux dérivées partielles, le gradient de la fonctionnelle entaché d'un bruit.

Aux § II.2.2 et II.3.1 on définit à partir des variables aléatoires $Z(x)$, des variables aléatoires Z_n utilisées à chaque itération et dont les propriétés se déduisent de celles de $Z(x)$ par les lois conditionnelles de Z_n à X_n fixé.

Cela signifie que dans les exemples, à chaque itération, on devra disposer d'une nouvelle mesure $\mathcal{E}_y(v_n) + \eta^{(n)}$ indépendante des précédentes.

On peut alors appliquer les algorithmes développés aux § précédents :

$$\left| \begin{array}{l} \text{Soit } v_0 \text{ quelconque dans } \mathcal{U}_{ad} \\ \text{Supposons } V_n \text{ calculé. Alors} \\ V_{n+1} = P_{\mathcal{U}_{ad}} [V_n + \rho_n (\nabla J(V_n) + Z_n)] \end{array} \right.$$

Pour calculer V_{n+1} , on aura besoin d'une réalisation de $\nabla J(V_n) + Z_n$.

Soit $v_n = V_n(\omega)$; l'observation (c'est-à-dire une mesure) fournit

$\mathcal{E}_y(v_n) + \eta$. On cherche alors $p_1(v_n)$ solution de (2.5.1) et (2.5.2) permet le calcul d'une réalisation de $\nabla J(v_n) + \xi$ qui est aussi une réalisation de $\nabla J(V_n) + Z_n$.

II.5.2.- Problème de contrôle dans un système parabolique

On reprend le problème et les notations définis au § I.1.2.

On sait (J.L. Lions [1]) que $\nabla J(v) = 2(B^* p(v) + \Lambda_{\mathcal{U}} N v)$

où $p(v)$ est solution du système adjoint :

$$\left\{ \begin{array}{l} - \frac{dp}{dt}(v) + A^*(t) p(v) = \mathcal{E}^* \Lambda_{\mathcal{H}} (\mathcal{E}_y(v) - z_d) \text{ sur }]0, T[\\ p(T; v) = 0 \\ p(v) \in L^2(0, T; V) \end{array} \right.$$

Avec l'observation⁽¹⁾ $\mathcal{E}_y(v) + \eta$, on résoud le système

$$\begin{cases} -\frac{dp_1(v)}{dt} + A^*(t) p_1(v) = \mathcal{E}^*_{\lambda_{\mathcal{E}}}(\mathcal{E}_y(v) + \eta - z_d) \\ p_1(T;v) = 0 \\ p_1(v) \in L^2(0,T;V) \text{ pour chaque } \eta(\omega) \end{cases}$$

ce qui permet de déduire comme au § II.5.1

$$\nabla J(v) + \xi$$

et d'utiliser de même les algorithmes d'optimisation stochastique.

On traiterait de façon analogue des problèmes de contrôle avec observation finale (cf. J.L. Lions [1]).

II.5.3.- Problème d'identification d'un système aux dérivées partielles

On reprend le problème et les notations définis au § I.1.3.

On sait, (G. Chavent [1]) que l'introduction d'un état adjoint permet encore de calculer le gradient de la fonctionnelle. En effet, soit ψ l'application $(v,y) \mapsto A(v)y$.

$$\text{On a : } \frac{\partial \psi}{\partial y}(v,y) \delta y + \frac{\partial \psi}{\partial v}(v,y) \delta v = 0$$

$\frac{\partial \psi}{\partial y} = A(v)$. Mais le calcul explicite de $\frac{\partial \psi}{\partial v}$ sera fait seulement dans chaque exemple.

$$\text{Soit le système adjoint : (2.5.3) } A(v)^* p = -2 \mathcal{E}^*_{\lambda_{\mathcal{E}}}(\mathcal{E}_y(v) - z_d)$$

$$\text{G. Chavent montre : (2.5.4) } \nabla J(v) \circ \delta v = (\delta v, (\frac{\partial \psi}{\partial v}(v,y))^* p)$$

ce qui justifie l'utilisation de l'état adjoint p pour le calcul de $\nabla J(v)$.

Pour expliciter davantage, précisons la forme de $A(v)$.

$$\text{Soit l'exemple : } v = a(x) \text{ et } A(v) = - \sum_{i=1}^n \frac{\partial}{\partial x_i} (a(x) \frac{\partial}{\partial x_i})$$

$$A(v)y = - \sum_{i=1}^n \frac{\partial}{\partial x_i} (a(x) \frac{\partial y}{\partial x_i})$$

$$\text{et } \frac{\partial \psi}{\partial v}(v,y) \circ \partial v = - \sum_{i=1}^n \frac{\partial}{\partial x_i} (\delta v \frac{\partial y}{\partial x_i})$$

(1) Rappelons que $\mathcal{E}$ est l'opérateur d'observation : observation répartie, observation ponctuelle, observation finale par exemple.

D'où $(\delta v, \nabla J(v)) = (\delta v, \sum_{i=1}^n \frac{\partial y}{\partial x_i} \frac{\partial p}{\partial x_i})$.

Avec l'observation bruitée $\mathcal{E}_y(v) + \eta$,

soit p_1 solution de (2.5.5) $A^*(v) p_1 = -2 \mathcal{E}^*_{\Lambda_{\mathcal{H}}}(\mathcal{E}_y(v) + \eta - z_d)$

$$(2.5.6) \quad \sum_{i=1}^n \frac{\partial y}{\partial x_i} \frac{\partial p_1}{\partial x_i} = \nabla J(v) + \xi$$

où ξ est une variable aléatoire dont les propriétés se déduisent de celles de η (cf. Jocelyne Rouyer [1]).

Par exemple $E(\eta) = 0 \implies E(\xi) = 0$.

On peut alors appliquer les algorithmes d'optimisation stochastique. Ici, il n'y a pas lieu de faire intervenir un opérateur de projection :

$$V_{n+1} = V_n - \rho_n (\nabla J(V_n) + Z_n) .$$

Soit $v_n = V_n(\omega)$. L'observation fournit $\mathcal{E}_y(v_n) + \eta$.

On cherche $p_1(v_n)$ solution de (2.5.5)

et (2.5.6) permet de calculer une réalisation de $\nabla J(v_n) + \xi$ qui est aussi une réalisation de $\nabla J(V_n) + Z_n$.

II.6.- Méthodes stochastiques de pénalisation

II.6.1.- Généralités

Les méthodes envisagées jusqu'ici nécessitent de savoir calculer la projection sur le domaine des contraintes. Dans le cas d'un convexe quelconque, cette question peut être difficile à résoudre. Par ailleurs, on a vu au § II.2.2 que les hypothèses stochastiques ont une forme particulièrement adaptée aux problèmes réels lorsque le problème d'optimisation est sans contrainte ou lorsque l'opérateur de projection est linéaire.

Il y a donc lieu de rechercher d'autres méthodes stochastiques pour traiter les problèmes où les algorithmes précédents ne sont pas justifiés. Dans cette direction, on pourra consulter H.J. Kushner [2] et H.J. Kushner, T. Gavin [1] pour une version stochastique de la méthode des directions admissibles. Ce point de vue fournit théoriquement la méthode de descente la plus générale. Malheureusement, les auteurs n'ont pu prouver jusqu'ici la convergence pour des algorithmes qui font le meilleur usage des observations (cf. H.J. Kushner, T. Gavin [1], p. 351).

La convergence n'est donc prouvée que pour des méthodes très coûteuses.

Nous envisagerons au chapitre V des méthodes de dualité et dans ce paragraphe, des méthodes de pénalisation.

II.6.2.- Définition du problème

Soit $f : \Phi \longrightarrow \mathbb{R}$ où Φ est un espace de Hilbert séparable

$$g_i : \Phi \longrightarrow \mathbb{R} \quad i = 1 \dots m .$$

Soit le problème $P : \begin{array}{l} \text{minimiser } f(x) . \\ g_i(x) \geq 0 \\ i = 1 \dots m \end{array}$

On suppose que f vérifie $\underline{H}_1, \underline{H}_2, \underline{H}_3, \underline{H}_4, \lambda > 0,$

et que $-g_i$ est convexe, continûment différentiable $i = 1 \dots m$.

On remplace le problème P par le problème sans contraintes P_ϵ appelé problème pénalisé :

$$P_\epsilon : \underset{x}{\text{minimiser}} \quad f(x) + \frac{1}{\epsilon} g(x) = \underset{x}{\text{minimiser}} \quad F^\epsilon(x)$$

$$\text{où } \epsilon > 0, \quad g(x) = \sum_{i=1}^m H_i g_i^2(x) \quad H_i = \begin{cases} 1 & \text{si } g_i(x) < 0 \\ 0 & \text{si } g_i(x) \geq 0 \end{cases} .$$

Dans ces conditions, g est convexe (P. Kenneth, M. Sibony, J.P. Yvon [1])

et continûment différentiable : $\nabla [H_i g_i^2(x)] = \begin{cases} 2 g_i(x) \nabla g_i(x) & \text{si } g_i(x) < 0 \\ 0 & \text{si } g_i(x) \geq 0 \end{cases}$

Le Problème P_ϵ admet dans ces conditions une solution unique x_ϵ et $x_\epsilon \xrightarrow{\epsilon \rightarrow 0} x$ solution unique de P dans Φ fort, (cf. P. Kenneth, M. Sibony, J.P. Yvon [1]).

Remarque : On a considéré une méthode extérieure écrite dans le cas de contraintes d'inégalités.

Les algorithmes qu'on va décrire maintenant pourraient s'adapter à des problèmes comprenant des contraintes d'égalité, à des variantes de la méthode extérieure, à des méthodes intérieures.

Nous allons à présent envisager des méthodes stochastiques pour résoudre le problème P_ϵ .

II.6.3. Algorithmes stochastiques, vérification des hypothèses

Nous supposons qu'en un point x , nous savons calculer exactement les contraintes. Nous connaissons donc $\nabla g(x)$.

L'observation nous fournit $\nabla f(x) + Z(x)$

où $Z(x)$ vérifie $\underline{H_5}$ (cas sans contrainte) et $\underline{H_6}$.

Nous connaissons donc $\nabla F^E(x) + Z(x)$.

Pour résoudre le problème P_E , nous allons donc utiliser dans le cas sans contrainte les méthodes développées au début de ce chapitre.

Les hypothèses sur le bruit : $\underline{H_5}$ $E(Z(x)) = 0$

$$\underline{H_6} \quad E(\|Z(x)\|^2) \leq B^2$$

sont légitimes. Il reste à s'assurer que $F^E(x)$ vérifie $\underline{H_1}$, $\underline{H_2}$, $\underline{H_3}$, $\underline{H_4}$ si f vérifie les mêmes hypothèses.

$\underline{H_1}$ et $\underline{H_2}$ sont satisfaites.

$\underline{H_4}$ est vérifiée avec le même $\lambda > 0$ que pour $f(x)$.

En effet g est convexe.

Donc $g(x) - g(y) \geq (\nabla g(y), x-y)$ (J. Céa [1])

et donc aussi $g(y) - g(x) \geq (\nabla g(x), y-x)$.

D'où $(\nabla g(x) - \nabla g(y), x-y) \geq 0$.

Il reste à vérifier $\underline{H_3}$ que les hypothèses faites sur g_i $i = 1 \dots m$ sont insuffisantes à assurer.

$$\underline{H_i} \quad g_i^2(x) = \begin{cases} 0 & \text{si la contrainte } g_i(x) > 0 \text{ est satisfaite} \\ g_i^2(x) & \text{sinon} \end{cases}$$

Nous supposons que hors du convexe $\{x \mid g_i(x) \geq 0\}$, la croissance de $g_i^2(x)$ n'est pas trop rapide lorsque $\|x\| \rightarrow +\infty$, ce qu'on exprime par la condition :

$$\underline{H'_4} : \quad \|\nabla g_i^2(x) - \nabla g_i^2(y)\| \leq \Lambda_i \|x-y\| \quad \text{si } g_i(x) < 0, \quad g_i(y) < 0.$$

Alors $\|\nabla [H_i g_i^2(x)] - \nabla [H_i g_i^2(y)]\| \leq \Lambda_i \|x-y\| \quad \forall x, y$

et $\|\nabla F^E(x) - \nabla F^E(y)\| \leq \left[\Lambda + \frac{1}{\epsilon} \sum_{i=1}^n \Lambda_i \right] \|x-y\| = \Lambda'(\epsilon) \|x-y\|$.

Remarque : S'il existe i tel que $\underline{H'_4}$ ne soit pas vérifiée, on peut remplacer $g(x)$ par $\sum_{i=1}^n H_i [-g_i(x)]^{1+\alpha}$, $\alpha \geq 0$.

Si $\alpha < 1$, on obtiendra une fonction à croissance plus lente et H_4' sera plus aisément satisfaite. En particulier si $\alpha = 0$, il suffit que ∇g_i satisfasse à une condition de Lipschitz. Notons que dans ce cas, $g(x)$ est toujours différentiable mais peut être pas continûment, ce qui est compatible avec les algorithmes envisagés.

On peut donc utiliser un algorithme du type

$$X_{n+1}^\varepsilon = X_n^\varepsilon - \rho_n (\nabla F^\varepsilon(X_n) + Z_n) .$$

On va envisager les deux choix de ρ_n :

$$\begin{aligned} - \rho_n > 0 \quad \sum_n \rho_n = +\infty \quad \sum_n \rho_n^2 < +\infty \\ - 0 < \delta(\varepsilon) \leq \rho_n \leq \frac{2\lambda}{\Lambda'^2(\varepsilon)} - \delta(\varepsilon) \quad (\implies \delta(\varepsilon) \leq \frac{\lambda}{\Lambda'^2(\varepsilon)}) . \end{aligned}$$

II.6.4.- Convergence

Algorithme de série divergente :

Proposition 2.6.4.1.-

$$\lim_{\varepsilon \rightarrow 0} \lim_{n \rightarrow +\infty} \|X_n^\varepsilon - x\|_{L^2}^2 = 0$$

En effet $\|X_n^\varepsilon - x\|_{L^2}^2 \leq 2 \left[\|X_n^\varepsilon - x_\varepsilon\|_{L^2}^2 + \|x_\varepsilon - x\|_\Phi^2 \right]$

$$\|x_\varepsilon - x\|_\Phi^2 \xrightarrow{\varepsilon \rightarrow 0} 0$$

et $\|X_n^\varepsilon - x_\varepsilon\|_{L^2}^2 \xrightarrow[n \rightarrow \infty]{} 0 \quad (\text{Th. 2.2.2.2}) .$

Algorithme du gradient

Proposition 2.6.4.2.-

$$\limsup_n \|X_n^\varepsilon - x\|_{L^2}^2 \xrightarrow{\varepsilon \rightarrow 0} 0$$

Démonstration :

On applique le théorème 2.3.2.4 en prenant $\rho_n \leq \frac{\lambda}{\Lambda'^2(\varepsilon)}$ à partir d'un certain rang.

$$\delta(\varepsilon) \leq \rho_n \leq \frac{\lambda}{\Lambda'^2(\varepsilon)} \quad \delta(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 0, \quad \frac{\lambda}{\Lambda'^2(\varepsilon)} \xrightarrow{\varepsilon \rightarrow 0} 0$$

$$\|X_n^\varepsilon - x\|_{L^2}^2 \leq 2 \left[\|X_n^\varepsilon - x_\varepsilon\|_{L^2}^2 + \|x_\varepsilon - x\|_\Phi^2 \right]$$

$$\limsup_n \|X_n^\varepsilon - x\|_{L^2}^2 \leq 2 \left[\frac{B^2}{\Lambda'^2(\varepsilon)} + \|x_\varepsilon - x\|_\Phi^2 \right] \xrightarrow{\varepsilon \rightarrow 0} 0.$$

3

optimisation stochastique

méthodes de gradient

l'observation est une

fonctionnelle aléatoire linéaire

Chapitre III

Optimisation stochastique - Méthodes de gradient

L'observation est une fonctionnelle aléatoire linéaire

III.1.- Problème

III.1.1.- Orientation

Les perturbations dues à des phénomènes non cohérents sont très bien représentées par des "bruits blancs", c'est-à-dire, dans le cas d'un processus discret, un processus gaussien discret $\xi(t_i)$ vérifiant :

$$\begin{aligned} E(\xi(t_i)) &= 0 \\ E(\xi(t_i) \xi(t_j)) &= \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \end{aligned}$$

Donc ce processus est formé de variables aléatoires gaussiennes centrées indépendantes.

La définition d'un "bruit blanc continu" conservant les propriétés intéressantes du bruit blanc discret soulève un certain nombre de difficultés (cf. A. Bensoussan [1]) qui nécessitent un contexte théorique plus élaboré que celui des variables aléatoires.

Une manière de définir un "bruit blanc continu" consiste à utiliser le concept de fonctionnelle aléatoire linéaire (1) :

Définition : Soit $\Phi = L^2(0,T)$, soit $\chi_\varphi(\omega)$ une F.A.L. telle que $(\chi_{\varphi_1}, \chi_{\varphi_2}, \dots, \chi_{\varphi_n})$ est un vecteur aléatoire gaussien de $\mathbb{R}^n \forall n$, $\forall \varphi_1, \dots, \varphi_n \in \Phi$ et telle que

$$E(\chi_\varphi) = 0 \quad \forall \varphi \in \Phi \quad (\text{i.e. } E(\chi) = 0)$$

$$E(\chi_{\varphi_1} \chi_{\varphi_2}) = (\varphi_1, \varphi_2) \quad \forall \varphi_1, \varphi_2 \in \Phi \quad (\text{i.e. } \text{cov}(\chi) = I)$$

On dit que χ_φ est un bruit blanc sur $[0,T]$.

III.1.2.- Le problème stochastique

Soit Φ un espace de Hilbert séparable.

On veut minimiser $f(x)$ sur K convexe fermé non vide $\subset \Phi$. Mais l'observation ne se réduit plus à une variable aléatoire.

(1) Une autre manière est d'utiliser la théorie des intégrales stochastiques.

On observe une réalisation d'une fonctionnelle aléatoire linéaire i.e. pour tout $\varphi \in \Phi$ et tout $x \in \Phi$, on connaît une réalisation de la variable aléatoire réelle $Y_\varphi(x) = (\nabla f(x), \varphi)_\Phi + Z_\varphi(x)$ où $E(Z_\varphi(x)) = 0$.

$Z_\cdot(x)$ est une F.A.L.

On ne peut définir un algorithme à l'aide de $Y_n = \nabla f(X_n) + Z_n$ car Z_n (et donc Y_n) ne serait pas une variable aléatoire, mais seulement une F.A.L. Par suite $f(X_n)$ n'aurait pas de sens.

III.2.- Problème approché

III.2.1.- Définition de Φ_h

Si Φ était de dimension finie, $Z_\cdot(x)$ se réduirait à une variable aléatoire à valeurs dans Φ car, sur un espace de dimension finie, toute application linéaire est nécessairement continue.

On va donc considérer une famille $\{\Phi_h\}$, $\Phi_h \subset \Phi$ de sous-espaces vectoriels fermés de dimension $N(h) < +\infty$.

Naturellement, les Φ_h ne seront pas choisis quelconques.

On va faire des hypothèses assurant qu'ils permettent de "bien représenter" K lorsque $h \rightarrow 0$.

Posons $K_h = K \cap \Phi_h$. K_h est convexe fermé.

Hypothèses sur $\{\Phi_h\}$:

- $h \geq h' \implies \Phi_h \subset \Phi_{h'}$,
- $\exists h (> 0)$ tel que $K_h \neq \emptyset$
- $\overline{\bigcup_{h>0} K_h} = K$.

Donnons un moyen de construction de $\{\Phi_h\}$ (méthode de Galerkin)

1) il existe une suite $\{k_n\}$ dense dans K :

En effet Φ est séparable ; il existe une suite $\{x_n\}$ dense dans Φ .

Soit $k \in K (\neq \emptyset)$.

Il existe une suite extraite $\{x_{n_v}\}$ telle que $x_{n_v} \rightarrow k$.

Par continuité $P x_{n_v} \rightarrow Pk = k$.

Donc la suite $\{k_n\}$ $k_n = P x_n$ est dense dans K .

2) Soit K_n = enveloppe convexe $(k_1, \dots, k_n)$

et Φ_n = sous-espace vectoriel fermé de Φ engendré par $(k_1, \dots, k_n)$.

On a alors des Φ_n croissants

et $\overline{\bigcup_n K_n} = K$ à cause de 1)

Notons π_h l'opérateur de projection sur Φ_h ,

P_h l'opérateur de projection sur K_h

et P l'opérateur de projection sur K .

On veut utiliser le

Théorème (Bourbaki [1] p. 134)

Soit Ψ un ensemble filtrant croissant de parties convexes, complètes et non vides de E , contenues dans une partie convexe, séparée et complète A de E . Si N est l'enveloppe fermée convexe des ensembles $H \in \Psi$, la projection d'un point $x \in E$ sur H tend vers la projection de x sur N suivant le filtre des sections de Ψ .

En appliquant ce résultat, compte-tenu des hypothèses faites sur $\{\Phi_h\}$, on a le

Lemme 3.2.1.-

$$\boxed{P_h u \xrightarrow{h \rightarrow 0} Pu \text{ dans } \Phi \text{ fort}}$$

Remarque : Si on traite un problème d'optimisation sans contrainte (minimisation sur Φ) , on peut faire sur $\{\Phi_h\}$ la seule hypothèse : $\pi_h u \xrightarrow{h \rightarrow 0} u$ dans Φ fort.

On n'a plus besoin de $\Phi_h \subset \Phi_h$, pour $h \geq h'$.

III.2.2.- Restriction de l'observation

Soit h fixé.

L'observation assure la connaissance de $Y_\varphi(x) \quad \forall \varphi \in \Phi, \quad \forall x \in \Phi$.

On restreint cette observation à la connaissance de $Y_\varphi(x) \quad \forall \varphi \in \Phi_h, \quad \forall x \in \Phi_h$, c'est-à-dire qu'on n'utilise pas toute l'information.

$$Y_\varphi(x) = (\nabla f(x), \varphi)_\Phi + Z_\varphi(x) \quad , \quad E(Z_\varphi(x)) = 0 \quad .$$

Puisque $\dim \Phi_h = N(h) < +\infty$,

l'application $\varphi \xrightarrow{Z \cdot (x)(\omega)} Z \varphi(x)(\omega)$, linéaire est aussi continue
 $\Phi_h \longrightarrow \mathcal{R}$

et d'après le théorème de Riesz, il existe $Z^h(x)(\omega) \in \Phi_h$ tel que

$$Z \varphi(x)(\omega) = (Z^h(x)(\omega), \varphi)_{\Phi}.$$

La connaissance de $(\nabla f(x), \varphi)_{\Phi} \quad \forall \varphi \in \Phi_h$ équivaut à la connaissance de $\pi_h \nabla f(x)$.

La restriction de l'observation à Φ_h conduit donc à la connaissance de la variable aléatoire à valeurs dans Φ_h :

$Y^h(x) = \pi_h \nabla f(x) + Z^h(x) \quad \forall x \in \Phi_h$, ce qui est l'observation pratique.

III.2.3.- Problème d'optimisation sur K_h .

On veut appliquer les algorithmes et les résultats de convergence du chap. II à un problème d'optimisation sur K_h convexe fermé de Φ_h . Cependant, à priori, comme $Y^h(x) = \pi_h \nabla f(x) + Z^h(x)$, on ne peut minimiser $f(x)$ sur K_h , mais seulement $f_h(x) = f(0) + \int_0^1 (\pi_h \nabla f(\theta x), x)_{\Phi} d\theta$.

En fait, puisque Φ_h est un sous-espace vectoriel fermé de Φ ,

$$(\pi_h \nabla f(\theta x), x)_{\Phi} = (\nabla f(\theta x), \pi_h x)_{\Phi}.$$

Donc si $x \in \Phi_h$, $f_h(x) = f(x)$.

Donc, minimiser $f_h(x)$ sur K_h est en fait minimiser f .

III.2.4.- Hypothèses

Il faut à présent se demander si les hypothèses des théorèmes (2.2.2.1), (2.2.2.2), (2.3.2.1), (2.3.2.2) sont satisfaites sur K_h .

Hypothèses sur f

On fait évidemment dans K , les hypothèses $\underline{H_1}$, $\underline{H_2}$, $\underline{H_3}$, $\underline{H_4}$
 f vérifie $\underline{H_1}$ et $\underline{H_2}$ sur K_h puisque $K_h \subset K$.

$\pi_h \nabla f(x)$ vérifie $\underline{H_3}$ et $\underline{H_4}$ sur K_h car :

$$\|\pi_h \nabla f(x) - \pi_h \nabla f(y)\|_{\Phi} = \|\pi_h(\nabla f(x) - \nabla f(y))\|_{\Phi} \leq \|\nabla f(x) - \nabla f(y)\| \leq \Lambda \|x-y\|_{\Phi} \quad (\underline{H}_3 \text{ sur } K)$$

$$(\pi_h \nabla f(x) - \pi_h \nabla f(y), x-y)_{\Phi} = (\nabla f(x) - \nabla f(y), \pi_h(x-y))_{\Phi} = (\nabla f(x) - \nabla f(y), x-y)_{\Phi} \geq \lambda \|x-y\|_{\Phi}^2 \quad (x, y \in K_h) \quad (\underline{H}_4 \text{ sur } K)$$

Hypothèses sur le bruit

a) algorithme de séries divergentes : on remplace $\underline{H}_5$ et $\underline{H}_6$ par

$$\underline{H}_5'' : E \left[P_h(x - \rho(\pi_h \nabla f(x) + Z^h(x))) \right] = P_h(x - \rho \pi_h \nabla f(x))$$

$$\underline{H}_6'' : E \left[\|Z^h(x)\|_{\Phi}^2 \right] \leq B_h^2 .$$

Remarques : 1 - Les remarques faites au § II.2.2 pour $\underline{H}_5$ et $\underline{H}_6$ restent valables pour $\underline{H}_5''$ et $\underline{H}_6''$.

2 - Etant données les propriétés attendues au § III.1.1 des F.A.L. considérées on ne peut espérer dans $\underline{H}_6'$ une constante indépendante de h .

b) algorithme de Gradient : on remplace $\underline{H}_5'$ et $\underline{H}_6'$ par

$$\underline{H}_5''' : E(Z^h(x)) = 0$$

$$\underline{H}_6''' : \exists H > 0 \text{ tel que}$$

$$h \leq H \implies E \left[(Z^h(x), P_h x - P_h(x - \rho(\pi_h \nabla f(x) + Z^h(x))))_{\Phi} \right] \leq B_h E \left[(\pi_h \nabla f(x), P_h x - P_h(x - \rho(\pi_h \nabla f(x) + Z^h(x))))_{\Phi} \right] .$$

On conserve les hypothèses $\underline{H}_7$ ou (selon le cas) $\underline{H}_7'$.

Remarque : Les remarques faites au § II.3.1 pour $\underline{H}_6'$, $\underline{H}_7$, $\underline{H}_7'$ restent valables ici.

III.2.5.- Algorithme et convergence

On considère donc l'algorithme

$$X_{n+1}^h = P_h (X_n^h - \rho_n Y_n^h) = P_h (X_n^h - \rho_n (\pi_h \nabla f(X_n^h) + Z_n^h))$$

où Z_n^h est une variable aléatoire à valeurs dans Φ_h dont la loi à $X_n^h = x$ fixé est identique à celle de $Z^h(x)$.

Pour l'algorithme de séries divergentes, ρ_n est tel que :

$$\rho_n > 0, \quad \sum_n \rho_n = +\infty, \quad \sum_n \rho_n^2 < +\infty.$$

Pour l'algorithme de Gradient, ρ_n est tel que :

$$\delta \leq \rho_n \leq \frac{2\lambda}{\Lambda^2} - \delta.$$

$$\text{En posant } L_h = \inf_{x \in K_h} f(x),$$

sous les hypothèses $\underline{H}_1, \dots, \underline{H}_4, \underline{H}_5'', \underline{H}_6'', \lambda > 0$ (resp. $\underline{H}_1, \dots, \underline{H}_4, \underline{H}_5''', \underline{H}_6''', H_7', \lambda > 0$)

Le théorème (2.2.2.2) (resp. le théorème (2.3.2.2)) assure :

- $\{z \in K_h \mid f(z) = L_h\}$ est réduit à un point, soit x_h
- $x_n^h \xrightarrow[n \rightarrow \infty]{} x_h$ presque sûrement et en moyenne quadratique

III.3.- Convergence de la solution du problème approché

Les hypothèses faites assurent : il existe $x \in K$ unique tel que $f(x) = L$.

Montrons le :

Théorème 3.3.1.-

$$x_h \xrightarrow[h \rightarrow 0]{} x \text{ dans } \Phi \text{ fort.}$$

On établira d'abord la convergence faible puis la convergence forte après avoir démontré 2 lemmes :

Lemme 3.3.1.-

$$L_h \xrightarrow[h \rightarrow 0]{} L$$

Démonstration :

$K_h \subset K \implies L_h \geq L$. Soit x le point unique de K tel que $f(x) = L$
 $P_h x \in K_h$. Donc $L \leq L_h \leq f(P_h x) \xrightarrow{} L$ (lemme 3.2.1 et f continue).

Lemme 3.3.2.-

Sous les hypothèses H_3 , H_4 , $\lambda > 0$,
 $\{x \in \Phi \mid f(x) \leq f(x_0)\}$ est borné dans Φ

Démonstration :

Soit x tel que $f(x) \leq f(x_0)$

$$f(x_0) - f(x) = -(\nabla f(x_0), x - x_0)_\Phi - \int_0^1 d\theta \underbrace{(\nabla f(x_0 + \theta(x - x_0)) - \nabla f(x_0), x - x_0)_\Phi}_{\geq 0 \quad (H_4)}$$

$$\implies -(\nabla f(x_0), x - x_0)_\Phi \geq 0$$

$$f(x_0) - f(x) = \underbrace{-(\nabla f(x_0), x - x_0)_\Phi}_{\geq 0} - \frac{\nu}{2} \|x - x_0\|^2$$

où $0 < \lambda \leq \nu \leq \Lambda$ (analogue à démonstration du théorème 2.2.2.1)

Si $\{x \mid f(x) \leq f(x_0)\}$ n'était pas borné, il existerait une suite $\{x_k\}$ telle que le 2ème membre $\xrightarrow[k \rightarrow \infty]{} -\infty$, en contradiction avec $f(x_0) - f(x_k) \geq 0$.

Démonstration du théorème 3.3.1 :

1 - $\{x_k\}$ est borné dans Φ :

$\exists x_0 \in K$ tel que $f(x_0) > L$. Sinon $f(x) \equiv L$ et le problème de minimiser f n'a pas d'intérêt. Posons $\varepsilon = f(x_0) - L > 0$

$$L_h \longrightarrow L \implies \exists H \text{ tel que } h < H \implies f(x_h) = L_h \leq L + \varepsilon = f(x_0)$$

Donc $\{x_h \mid h < H\} \subset \{x \mid f(x) \leq f(x_0)\}$ borné dans Φ (lemme 3.3.2)).

2 - Il existe une suite extraite $\{x_{h_n}\}$: $x_{h_n} \longrightarrow x$ dans Φ faible

il existe une sous-suite $\{x_{h_n}\}$ convergente :

$$x_{h_n} \xrightarrow[n \rightarrow \infty]{} x' \text{ dans } \Phi \text{ faible, } h_n \xrightarrow[n \rightarrow \infty]{} 0$$

Montrons que $x' = x \iff f(x') = L$ (unicité de x)

$$\begin{array}{ccc} f(x') - L_{h_n} = f(x') - f(x_{h_n}) \leq -(\nabla f(x'), x_{h_n} - x')_\Phi & \text{comme précédemment} \\ \downarrow n \rightarrow \infty & \text{(lemme 3.3.1)} & \downarrow n \rightarrow \infty & \text{(convergence faible)} \\ f(x') - L & & 0 \end{array}$$

Donc $f(x') \leq L \quad x' \in K \implies f(x') = L \iff x' = x$.

3 - $x_h \xrightarrow{h \rightarrow 0} x$ dans Φ faible

En effet, la limite trouvée est indépendante de la suite extraite choisie.

4 - $x_h \xrightarrow{\quad} x$ dans Φ fort

Comme dans la démonstration du lemme 3.3.2

$$L - L_h = f(x) - f(x_h) = -(\nabla f(x), x_h - x)_\Phi - \frac{\nu}{2} \|x - x_h\|_\Phi^2 \quad (0 < \lambda \leq \nu \leq \Lambda)$$

$$\implies \frac{\lambda}{2} \|x - x_h\|_\Phi^2 \leq \underbrace{-(\nabla f(x), x_h - x)_\Phi - L + L_h}_{\substack{\downarrow h \rightarrow 0 \text{ (lemme 3.3.1 et convergence faible)} \\ 0}}$$

III.4.- Exemple d'application

Soit le problème d'identification déjà envisagé aux paragraphes I.1.3 et II.5.3.

Cette fois, l'observation est $\mathcal{E}_y(v) + \eta$ où η n'est plus une variable aléatoire, mais une F.A.L., par exemple un bruit blanc. (cf. A. Bensoussan [1]).

$\mathcal{E}_y(v)$ est une fonction si bien que l'observation est réalisée en pratique par la connaissance (entachée d'un bruit) de $\mathcal{E}_y(v)$ en un certain nombre de points choisis en fonction de la résolution par discrétisation des systèmes aux dérivées partielles concernés.

La restriction de l'observation aux points de discrétisation fournit $[\mathcal{E}_y(v) + \eta]_h$, vecteur de dimension finie ; on a donc remplacé la F.A.L. η par le vecteur aléatoire $[\eta]_h$ de dimension finie.

La mise en oeuvre de la méthode de l'état adjoint pour calculer le gradient (cf. § II.5.3) conduit à la résolution de deux systèmes aux dérivées partielles, le système direct et le système adjoint.

On effectue cette résolution sur ordinateur par discrétisation des systèmes où on n'utilise en fait que la restriction de l'observation indiquée, ce qui fournit une restriction à un sous-espace de dimension finie de la F.A.L. $\nabla J(v) + \xi$, soit $\pi_h(\nabla J(v) + \xi) = \pi_h \nabla J(v) + Z^h(v)$ où $Z^h(v)$ est une variable aléatoire.

L'algorithme sera donc : (problème sans contrainte)

$$V_{n+1}^h = V_n^h - \rho_n (\pi_h \nabla J(V_n^h) + Z_n^h)$$

où Z_n^h est une variable aléatoire à valeurs dans Φ_h dont la loi à $V_n^h = v$ fixé est identique à celle de $Z^h(v)$.

4

optimisation stochastique

méthodes de

gradient approché

IV.1.- Définition du problème

Soit Φ un espace de Hilbert sur $\mathbb{R}$ de dimension finie p et $f : \Phi \longrightarrow \mathbb{R}$.

On veut minimiser $f(x)$ sur K convexe fermé non vide de Φ .

Observation

On suppose à présent que l'observation fournit seulement des réalisations de la variable aléatoire réelle $Y(x) = f(x) + Z(x) \quad \forall x \in K, \quad E(Z(x)) = 0$.

(Dans le cas déterministe, l'observation serait $f(x)$).

L'idée générale est de calculer une valeur approchée du gradient en approchant par différences les dérivées partielles dans la direction des axes de coordonnées.

La restriction à Φ de dimension finie est donc fondamentale. L'algorithme étudié au § IV.2 est celui de Kiefer-Wolfowitz et a déjà fait l'objet de nombreuses publications (cf. par exemple Kiefer-Wolfowitz [1]). Mais les hypothèses faites ici sont différentes et plus conformes aux hypothèses classiques en théorie de l'optimisation. L'algorithme étudié au § IV.3 n'a à notre connaissance, été l'objet d'aucune publication. Nous en donnons donc la démonstration de la convergence même dans le cas déterministe.

IV.2.- Algorithme de série divergente stochastique

IV.2.1.- Algorithme, hypothèses

On conserve les hypothèses $\underline{H}_1$, $\underline{H}_3$, $\underline{H}_4$ (chap. II) et on remplace $\underline{H}_2$ par $\underline{H}'_2$: f est 2 fois différentiable sur K et ses dérivées y sont bornées jusqu'au 2ème ordre.

Les hypothèses stochastiques sont la traduction dans le cas présent des hypothèses du § II.2.2.

$\underline{H}_5$ est remplacé par :

$$\underline{H}_{15} \quad E [P(x - \rho \Delta(x, c))] = P(x - \rho \Delta f(x, c))$$

où $\Delta(x, c)$, $\Delta f(x, c)$ sont définis ci-dessous.

Puisque $Z(x)$ est une variable aléatoire réelle, $\underline{H}_6$ est remplacé par

$$\underline{H}_{16} \quad E [(Z(x))^2] \leq A$$

Φ étant rapporté à une base orthonormée $\{\vec{u}_i\}$ $i = 1 \dots p$

$$\begin{aligned} \text{on pose } \Delta(x, c) &= \sum_{i=1}^p \vec{u}_i \frac{Y(x + c\vec{u}_i) - Y(x - c\vec{u}_i)}{2c} \\ &= \sum_{i=1}^p \vec{u}_i \frac{f(x + c\vec{u}_i) - f(x - c\vec{u}_i)}{2c} + \sum_{i=1}^p \vec{u}_i \frac{Z(x + c\vec{u}_i) - Z(x - c\vec{u}_i)}{2c} \\ &= \Delta f(x, c) + \Delta Z(x, c) \end{aligned}$$

Algorithme : Soit P le projecteur sur K

$$\left| \begin{array}{l} X_0 \in L^2(\omega, \Phi) \text{ (par exemple un élément fixe de } K) \\ X_{n+1} = P(X_n - \rho_n \Delta_n) \end{array} \right.$$

$$\Delta_n = \sum_{i=1}^p \vec{u}_i \frac{Y(x + c_n \vec{u}_i) - Y(x - c_n \vec{u}_i)}{2c_n}$$

$$Y(x + c_n \vec{u}_i) = f(X_n + c_n \vec{u}_i) + Z(X_n + c_n \vec{u}_i)$$

où $Z(X_n + c_n \vec{u}_i)$ (resp. $Z(X_n - c_n \vec{u}_i)$) est une variable aléatoire réelle qui, à $X_n + c_n \vec{u}_i = x$ (resp. $X_n - c_n \vec{u}_i = x$) fixé, a même loi que $Z(x)$.

ρ_n et c_n satisfont aux conditions suivantes :

$$\begin{aligned} c_n, \rho_n &> 0, \quad \sum_n \rho_n = +\infty, \quad \sum_n \rho_n c_n < +\infty, \quad \sum_n \left(\frac{\rho_n}{c_n}\right)^2 < +\infty \\ c_n &\xrightarrow{n \rightarrow \infty} 0, \quad \rho_n \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

Ces conditions sont compatibles comme le montre le choix habituel :

$$\rho_n = \rho n^{-1}, \quad c_n = c n^{-\gamma} \quad \text{où } c, \rho > 0 \text{ et } 0 < \gamma < \frac{1}{2}.$$

Par ailleurs, le système de ces conditions n'est pas redondant (cf. B. Fichet [1]).

Remarques : 1 - L'algorithme défini ici est l'algorithme de Kiefer-Wolfowitz classique. Mais la convergence est assurée ici par des hypothèses usuelles en théorie déterministe de l'optimisation, ce qui est une approche relativement nouvelle. J.H. Venter [1] démontre un théorème relevant des mêmes préoccupations. Nous le comparons à nos résultats au § IV.4 ci-dessous.

2 - On a fait un abus de notation : $f(X_n + c_n \vec{u}_i)$ représente la valeur de la fonction f appliquée à $X_n + c_n \vec{u}_i$, ce qui n'est pas le cas pour $Z(X_n + c_n \vec{u}_i)$. $Z(x)$ et $Z(X_n + c_n \vec{u}_i)$, x déterministe, $X_n + c_n \vec{u}_i$ aléatoire, n'ont pas la même signification.

3 - Comme aux § II.2.2 et II.3.2, $X_n \in L^2(\underline{\omega}, \Phi)$.

IV.2.2.- Convergence de l'algorithme

Modifions l'écriture de Δ_n :

$$\frac{f(X_n + c_n \vec{u}_i) - f(X_n - c_n \vec{u}_i)}{2c_n} = \frac{\partial f}{\partial x_i}(X_n) + \frac{1}{4} c_n \{f''_{i,i}(X_n + c_n \phi'_{i,n} \vec{u}_i) - f''_{i,i}(X_n - c_n \phi''_{i,n} \vec{u}_i)\}$$

$$0 \leq \phi'_{i,n}, \phi''_{i,n} \leq 1$$

On peut donc écrire

$$\Delta_n = \nabla f(X_n) + L_n + \sum_{i=1}^p \vec{u}_i \frac{Z(X_n + c_n \vec{u}_i) - Z(X_n - c_n \vec{u}_i)}{2c_n}$$

$$L_n = \frac{c_n}{4} \sum_{i=1}^p \vec{u}_i \{f''_{i,i}(X_n + c_n \phi'_{i,n} \vec{u}_i) - f''_{i,i}(X_n - c_n \phi''_{i,n} \vec{u}_i)\}$$

$$\underline{H'_2} \implies \|L_n\|_{\Phi} \leq c_n K_2$$

Théorème 4.2.2.1.-

Si $\underline{H_1}$, $\underline{H'_2}$, $\underline{H_3}$, $\underline{H_4}$, $\underline{H_{15}}$, $\underline{H_{16}}$ sont vérifiées, $E[f(X_n)] \xrightarrow{n \rightarrow \infty} L = \inf_{x \in K} f(x)$
--

Etablissons d'abord :

Lemme 4.2.2.1.-

$E[f(x_n)] \xrightarrow{n \rightarrow \infty} L'$

Démonstration :

$$f(X_n) - f(X_{n+1}) = -(\nabla f(X_n), X_{n+1} - X_n)_{\Phi} - \frac{\nu_n}{2} \|X_{n+1} - X_n\|_{\Phi}^2$$

$$\lambda \leq \nu_n \leq \Lambda$$

Posons $A = \frac{B^2}{P}$

$$E(Z(X_n + c_n \vec{u}_1) - Z(X_n - c_n \vec{u}_1))^2 \leq 2 [E(Z(X_n + c_n \vec{u}_1))^2 + E(Z(X_n - c_n \vec{u}_1))^2] \\ \leq 4 \frac{B^2}{P} .$$

$$\text{Donc } -\rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-X_n \right)_{L^2} \leq B^2 \frac{\rho_n^2}{c_n^2}$$

$$\text{et } \left| \frac{\theta_n' v_n'}{2} \rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-X_n \right)_{L^2} \right| \leq \frac{A}{2} B^2 \frac{\rho_n^2}{c_n^2}$$

$$(L_n, X_{n+1}-X_n)_{L^2} = (L_n, P(X_n - \rho_n (\nabla f(X_n) + L_n)) - P X_n)_{L^2}$$

$$|(L_n, X_{n+1}-X_n)|_{L^2} \leq \underbrace{\|L_n\|_{L^2}}_{\leq c_n K_2} \rho_n \cdot \underbrace{\|\nabla f(X_n) + L_n\|_{L^2}}_{\leq K_3 \quad (H_2')}_{L^2}$$

$$|(L_n, X_{n+1}-X_n)|_{L^2} \leq K_4 \rho_n c_n$$

$$\text{puisque } \sum_n \left(\frac{\rho_n}{c_n} \right)^2 < +\infty, \quad \sum_n \rho_n c_n < +\infty,$$

$$\sum_n w_n < +\infty .$$

On peut donc utiliser la proposition 1.5.2.2.

Lemme 4.2.2.2.-

$-\rho_n (\nabla f(X_n) + L_n, X_{n+1}-X_n)_{L^2}$ et $-(\nabla f(X_n) + L_n, X_{n+1}-X_n)_{L^2}$
sont les termes généraux ≥ 0 de séries convergentes

Démonstration :

$$\sum_n v_n < +\infty \quad \text{et} \quad \left(\frac{1}{\rho_n} - \frac{v_n' \theta_n'}{2} \right) \geq 1 \quad \text{pour } n \geq N_1 .$$

Lemme 4.2.2.3.-

$\{X_n\}$ est borné dans $L^2(\underline{\omega}, \Phi)$

Démonstration :

$$\begin{aligned}
 \|X_{n+1} - X_n\|_{L^2}^2 &= -\theta'_n \rho_n (\nabla f(X_n) + L_n, X_{n+1} - X_n)_{L^2} \\
 &\quad - \theta'_n \rho_n \left(\sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1} - X_n \right)_{L^2} \\
 &\leq -\rho_n (\nabla f(X_n) + L_n, X_{n+1} - X_n)_{L^2} + B^2 \left(\frac{\rho_n}{c_n} \right)^2 \\
 \implies \sum_n \|X_{n+1} - X_n\|_{L^2}^2 &+ \infty .
 \end{aligned}$$

Démonstration du théorème 4.2.2.1 :

Il suffit de trouver une sous-suite $\{X_{n_k}\}$ telle que $E[f(X_{n_k})] \xrightarrow[k \rightarrow \infty]{} L$.

Soit $\alpha > L$ fixé et $\epsilon > 0$ tel que $\alpha > \alpha - 2\epsilon > L$.

Soit $\hat{S}_\alpha$ défini comme dans la démonstration du théorème 2.2.2.1.

Hypothèse (dont on veut démontrer l'absurdité)

$$\exists N (\geq N_0) \text{ tel que } n \geq N \implies X_n \notin \hat{S}_\alpha$$

$$X_n \notin \hat{S}_\alpha, X_n \in L^2(\underline{\omega}, \Phi), X_n(\omega) \in K \implies E[f(X_n)] > \alpha.$$

Soit $X(\omega) = \tilde{x} \quad \forall \omega$ où $\tilde{x} \in S_{\alpha-2\epsilon}$ ($S_{\alpha-2\epsilon}$ défini dans la démonstration du théorème 2.2.1.1)

$$f(X) - f(X_n) \geq (\nabla f(X_n), X - X_n)_\Phi \quad (f \text{ convexe})$$

$$- (\nabla f(X_n), X - X_n)_{L^2} \geq E[f(X_n)] - E[f(X)] \geq \alpha - (\alpha - 2\epsilon)$$

$$\text{ou } - (\nabla f(X_n), X - X_n)_{L^2} - 2\epsilon \geq 0$$

$$|(L_n, X - X_n)_{L^2}| \leq \|L_n\|_{L^2} \|X - X_n\|_{L^2} \leq \underbrace{K_2 c_n}_{\downarrow 0 \text{ } n \rightarrow \infty} \underbrace{\|X - X_n\|_{L^2}}_{\text{borné}} \quad (\text{lemme 4.2.2.3})$$

Donc $(L_n, X - X_n)_{L^2} \leq \epsilon$ pour n assez grand.

$$- (L_n, X - X_n)_{L^2} + \epsilon - (\nabla f(X_n), X - X_n) - 2\epsilon \geq (-\nabla f(X_n), X - X_n) - 2\epsilon \geq 0$$

$$\text{ou } (4.2.2) \quad - (\nabla f(X_n) + L_n, X - X_n)_{L^2} - \epsilon \geq 0.$$

D'après la proposition 1.5.1.2 :

$$\begin{aligned} \|X_{n+1} - \tilde{x}\|_{L^2}^2 &\leq \|X_n - \tilde{x}\|_{L^2}^2 - 2\rho_n (\nabla f(X_n) + L_n + \sum_{i=1}^p u_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1} - \tilde{x})_{L^2} \\ &\leq \|X_n - \tilde{x}\|_{L^2}^2 - 2\rho_n (\nabla f(X_n) + L_n, X_{n+1} - \tilde{x})_{L^2} - 2\rho_n (\sum_{i=1}^p u_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1} - \tilde{x})_{L^2} \\ &\text{d'après (4.2.2)} \\ &\quad - 2\rho_n (\nabla f(X_n) + L_n, \tilde{x} - X_n)_{L^2} - 2\rho_n \varepsilon \end{aligned}$$

$$(4.2.3) \quad \|X_{n+1} - \tilde{x}\|_{L^2}^2 \leq \|X_n - \tilde{x}\|_{L^2}^2 - 2\rho_n (\nabla f(X_n) + L_n, X_{n+1} - X_n)_{L^2} + 2 \frac{\rho_n^2}{c_n^2} B^2 - 2\rho_n \varepsilon.$$

En itérant (4.2.3)

$$\begin{aligned} \|X_{N+m+1} - \tilde{x}\|_{L^2}^2 &\leq \|X_N - \tilde{x}\|_{L^2}^2 + 2 \sum_{n=N}^{N+m} \left[-\rho_n (\nabla f(X_n) + L_n, X_{n+1} - X_n)_{L^2} + \frac{\rho_n^2}{c_n^2} B^2 \right] \\ &\quad - 2\varepsilon \sum_{n=N}^{N+m} \rho_n. \end{aligned}$$

Le 2ème terme de 2ème membre est ≥ 0 et la série converge (lemme 4.2.2.2)

$\sum_n \rho_n = +\infty \implies$ le 2ème membre est < 0 pour m assez grand d'où l'absurdité.

Théorème 4.2.2.2.-

Sous les mêmes hypothèses et si de plus $\lambda > 0$,
 $\{z \in K \mid f(z) = L\}$ est réduit à un point, soit x
 et $X_n \xrightarrow[n \rightarrow \infty]{} x$ en moyenne quadratique

Démonstration identique à théorème 2.2.2.2.

Théorème 4.2.2.3.-

Sous les mêmes hypothèses qu'au théorème 4.2.2.2
 $X_n \xrightarrow[n \rightarrow \infty]{} x$ presque sûrement

Démonstration :

D'après la proposition 1.5.1.2

$$\begin{aligned} \|X_{n+1}-x\|_{\Phi}^2 &\leq \|X_n-x-\rho_n(\nabla f(X_n)+L_n-\nabla f(x))\|_{\Phi}^2 - 2\rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-x\right)_{\Phi} \\ &\leq (1-2\rho_n\lambda+\rho_n^2\Lambda^2) \|X_n-x\|_{\Phi}^2 + \rho_n^2 \|L_n\|_{\Phi}^2 + 2\rho_n \|L_n\|_{\Phi} \|X_n-x\|_{\Phi} + \\ &\quad + 2\rho_n^2 \|L_n\|_{\Phi} \|\nabla f(X_n)-\nabla f(x)\|_{\Phi} - 2\rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-x\right)_{\Phi} \end{aligned}$$

$$\|L_n\|_{\Phi} \leq c_n K_2, \quad \|\nabla f(X_n) - \nabla f(x)\|_{\Phi} \leq A \quad (\underline{H'_2}) \implies$$

$$\begin{aligned} \|X_{n+1}-x\|_{\Phi}^2 &\leq (1-2\rho_n\lambda+\rho_n^2\Lambda^2) \|X_n-x\|_{\Phi}^2 + \rho_n c_n \left[\rho_n c_n K_2^2 + 2K_2 \|X_n-x\|_{\Phi} + 2\rho_n K_2 A \right] \\ &\quad - 2\rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-x\right)_{\Phi} \end{aligned}$$

$$1 + u^2 \geq 2u \implies$$

$$\begin{aligned} \|X_{n+1}-x\|_{\Phi}^2 &\leq (1 - 2\rho_n\lambda + \rho_n^2\Lambda^2 + K_2 \rho_n c_n) \|X_n-x\|_{\Phi}^2 \\ &\quad + \rho_n c_n \left[\rho_n c_n K_2^2 + 2K_2 + 2\rho_n K_2 A \right] \\ &\quad - 2\rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-x\right)_{\Phi} \end{aligned}$$

$$1 - 2\rho_n\lambda + \rho_n^2\Lambda^2 + K_2 \rho_n c_n \underset{\rho_n \rightarrow 0}{\sim} 1 - \rho_n(2\lambda - K_2 c_n)$$

$$c_n \xrightarrow{>} 0 \implies 0 \leq 1 - 2\rho_n\lambda + \rho_n^2\Lambda^2 + K_2 \rho_n c_n \leq 1 \quad \text{pour } n \text{ assez grand}$$

$$\|X_{n+1}-x\|_{\Phi}^2 \leq \|X_n-x\|_{\Phi}^2 + K_0 \rho_n c_n - 2\rho_n \left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-x\right)_{\Phi}$$

$$E \left[\|X_{n+1}-x\|_{\Phi}^2 | X_1 \dots X_n \right] \leq \|X_n-x\|_{\Phi}^2 + K_0 \rho_n c_n - 2\rho_n E \left[\left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-x\right)_{\Phi} | X_1 \dots X_n \right]$$

$$\leq \|X_n-x\|_{\Phi}^2 + K_0 \rho_n c_n - 2\rho_n E \left[\left(\sum_1^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1}-P(X_n-\rho_n(\nabla f(X_n)+L_n))\right)_{\Phi} | X_1 \dots X_n \right]$$

$$\leq \|X_n-x\|_{\Phi}^2 + K_0 \rho_n c_n + 2\rho_n^2 B^2 \quad (1)$$

(1) On restreint la généralité du bruit de façon analogue à ce qui a été fait au théorème 2.2.2.3.

On peut maintenant achever la démonstration comme celle du théorème 2.2.2.3 en remplaçant simplement $2\rho_n^2 B^2$ par $2\rho_n^2 B^2 + K_0 \rho_n c_n$.

IV.3.- Algorithme de Gradient

IV.3.1.- Algorithme déterministe

IV.3.1.1.- Algorithme, hypothèses

On fait comme au § IV.2.1 les hypothèses $\underline{H}_1$, $\underline{H}_2'$, $\underline{H}_3$, $\underline{H}_4$.

Algorithme :

$$X_{n+1} = P(X_n - \rho_n \Delta_n) \quad 0 < \delta \leq \rho_n \leq \frac{2}{\Lambda} - \delta$$

Φ étant rapporté à une base orthonormée $\{\vec{u}_i\}$ $i = 1 \dots p$

$$\Delta_n = \sum_{i=1}^p \vec{u}_i \frac{f(X_n + c_n \vec{u}_i) - f(X_n - c_n \vec{u}_i)}{2c_n} \quad \text{où } c_n > 0 \quad \sum_n c_n < +\infty.$$

IV.3.1.2.- Convergence de l'algorithme

On peut comme au § IV.2.2 modifier l'écriture de Δ_n : $\Delta_n = \nabla f(X_n) + L_n$

$$\text{où } L_n = \frac{c_n}{4} \sum_{i=1}^p \vec{u}_i \{f''_{i,i}(X_n + c_n \phi'_{i,n} \vec{u}_i) - f''_{i,i}(X_n - c_n \phi''_{i,n} \vec{u}_i)\}$$

$$0 \leq \phi'_{i,n}, \phi''_{i,n} \leq 1$$

$$\underline{H}_2' \implies (4.3.1) \quad \|L_n\|_{\Phi} \leq c_n K_2.$$

Théorème 4.3.1.1.-

Si $\underline{H}_1$, $\underline{H}_2'$, $\underline{H}_3$, $\underline{H}_4$ sont vérifiées $f(X_n) \xrightarrow{n \rightarrow \infty} L = \inf_{x \in K} f(x)$
--

Etablissons d'abord

Lemme 4.3.1.1.-

$f(X_n) \xrightarrow{n \rightarrow \infty} L'$
--

Démonstration :

$$(2.2.1) \text{ est valable : } f(X_n) - f(X_{n+1}) = -(f(X_n), X_{n+1} - X_n)_\Phi - \frac{v_n}{2} \|X_{n+1} - X_n\|_\Phi^2$$

$$0 \leq \lambda \leq v_n \leq \Lambda$$

$$\begin{aligned} -\rho_n (\nabla f(X_n), X_{n+1} - X_n) &= (X_n - \rho_n (\nabla f(X_n) + L_n) - X_n + \rho_n L_n, P(X_n - \rho_n (\nabla f(X_n) + L_n) - PX_n)_\Phi \\ &= \mu^2(\rho_n) \|X_{n+1} - X_n\|_\Phi^2 + \rho_n (L_n, X_{n+1} - X_n)_\Phi \end{aligned}$$

$$\mu^2(\rho_n) \geq 1 \quad \text{d'après (1.5.2)}$$

$$|(L_n, X_{n+1} - X_n)|_\Phi \leq \|L_n\|_\Phi \|X_{n+1} - X_n\|_\Phi = \|L_n\|_\Phi \|P(X_n - \rho_n (\nabla f(X_n) + L_n) - PX_n)\|_\Phi$$

$$\leq \|L_n\|_\Phi \|\nabla f(X_n) + L_n\|_\Phi \leq c_n K_1 \quad (1.5.3) \quad \underbrace{(H'_2)}$$

$$\text{Donc } f(X_n) - f(X_{n+1}) = \left[\frac{\mu^2(\rho_n)}{\rho_n} - \frac{v_n}{2} \right] \|X_{n+1} - X_n\|_\Phi^2 + (L_n, X_{n+1} - X_n)_\Phi$$

$$= v_n + w_n$$

$$\sum_n |w_n| < +\infty \quad \text{car} \quad \sum_n c_n < +\infty$$

$$v_n \geq 0 \quad \text{car} \quad \frac{\mu^2(\rho_n)}{\rho_n} - \frac{v_n}{2} \geq \frac{1}{\frac{\Lambda}{2} - \delta} - \frac{1}{\frac{\Lambda}{2}} > 0$$

On peut donc utiliser la proposition 1.5.2.2.

Lemme 4.3.1.2.-

$$\sum_n \mu^2(\rho_n) \|X_{n+1} - X_n\|_\Phi^2 < +\infty$$

et $\{X_n\}$ est borné dans Φ

Résulte directement de l'utilisation de la proposition 1.5.2.2.

Démonstration du théorème 4.3.1.1 :

Comme au § II.3.2, si $L' > L$ soit $z \in K$ tel que $f(z) < L'$

$f(z) - L' = -2\varepsilon$ ($\varepsilon > 0$). Soit $n \geq N(\varepsilon)$ tel que $-\varepsilon \leq f(X_n) - L' \leq \varepsilon$

$$0 > -\varepsilon \geq f(z) - f(X_n) \geq (\nabla f(X_n), z - X_n)_\Phi \quad (f \text{ convexe})$$

$$\begin{aligned}
 0 &> -\varepsilon\delta \geq (\rho_n \nabla f(X_n), z - X_n) = (\rho_n (\nabla f(X_n) + L_n), z - X_n) + (\rho_n L_n, X_n - z) \\
 &= (X_n - \rho_n (\nabla f(X_n) + L_n) - X_n, X_n - X_{n+1}) + (\rho_n (\nabla f(X_n) + L_n), z - X_{n+1}) + (\rho_n L_n, X_n - z) \\
 &= -(X_n - \rho_n (\nabla f(X_n) + L_n) - X_n, P(X_n - \rho_n (\nabla f(X_n) + L_n)) - P X_n) \\
 &\quad + (\rho_n (\nabla f(X_n) + L_n), z - X_{n+1}) + (\rho_n L_n, X_n - z) \\
 0 &> -\varepsilon\delta \geq -\mu^2(\rho_n) \|X_n - X_{n+1}\|_\Phi^2 + (\rho_n L_n, X_n - z) + (X_n - X_{n+1}, z - X_{n+1}) \\
 &\quad + (X_{n+1} - X_n + \rho_n (\nabla f(X_n) + L_n), z - X_{n+1})_\Phi .
 \end{aligned}$$

Le dernier terme est ≥ 0 (1.5.1)

$$\text{Donc (4.3.2) } 0 > -\varepsilon\delta > -\mu^2(\rho_n) \|X_n - X_{n+1}\|_\Phi^2 + (\rho_n L_n, X_n - z) + (X_n - X_{n+1}, z - X_{n+1})_\Phi .$$

D'après le lemme 4.3.1.2 et (4.3.1) la relation (4.3.2) est absurde.

Donc $L' = L$.

Théorème 4.3.1.2.-

Si $\underline{H_1}, \underline{H'_2}, \underline{H_3}, \underline{H_4}, \lambda > 0, \delta \leq \rho_n < \frac{2\lambda}{\Lambda^2} - \delta$,
 . $\{z \in K \mid f(z) = L\}$ est réduit à un point soit x
 . $X_n \xrightarrow[n \rightarrow \infty]{} x$ dans fort

Démonstration :

On sait que sous ces hypothèses, x existe et est unique.

Reste à montrer : $X_n \xrightarrow[n \rightarrow \infty]{} x$.

On sait qu'alors $x = P(x - \rho_n \nabla f(x)) \quad \forall \rho_n : \delta \leq \rho_n \leq \frac{2\lambda}{\Lambda^2} - \delta \leq \frac{2}{\Lambda} - \delta$

$$X_{n+1} = P(X_n - \rho_n (\nabla f(X_n) + L_n)) .$$

$$\begin{aligned}
 \text{D'où (4.3.3) } \|X_{n+1} - x\|^2 &\leq \|X_n - x - \rho_n (\nabla f(X_n) - \nabla f(x)) - \rho_n L_n\|^2 \\
 &\leq (1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) \|X_n - x\|^2 + \rho_n^2 \|L_n\|^2 + 2\rho_n \|L_n\| \|X_n - x\| \\
 &\quad + 2\rho_n^2 \|L_n\| \|\nabla f(X_n) - \nabla f(x)\| .
 \end{aligned}$$

$$\max_{\delta \leq \rho \leq \frac{2\lambda}{\Lambda^2} - \delta} (1 - 2\rho\lambda + \rho^2 \Lambda^2) = 1 - 2\delta\lambda + \delta^2 \Lambda^2 = 1 - \alpha < 1 \quad \text{pour } \delta < \frac{2\lambda}{\Lambda^2} \quad \text{ce qui est vrai}$$

$$\|X_{n+1} - x\|^2 \leq (1-\alpha) \|X_n - x\|^2 + 2\rho_n K_2 c_n \|X_n - x\| + 2\rho_n^2 K_2 c_n \Lambda \|X_n - x\| + \rho_n^2 K_2^2 c_n^2$$

$$1 + u^2 \geq 2u \implies$$

$$\|X_{n+1} - x\|^2 \leq (1-\alpha) \|X_n - x\|^2 + \rho_n K_2 c_n (1 + \rho_n \Lambda) (1 + \|X_n - x\|^2) + \rho_n^2 K_2^2 c_n^2$$

$$\begin{aligned} \|X_{n+1} - x\|^2 &\leq (1-\alpha + (\frac{2\lambda}{\Lambda^2} - \delta)(1 + \frac{2\lambda}{\Lambda} - \delta\Lambda) K_2 c_n) \|X_n - x\|^2 \\ &\quad + (\frac{2\lambda}{\Lambda^2} - \delta) K_2 c_n \left[1 + \frac{2\lambda}{\Lambda} - \delta\Lambda + K_2 c_n (\frac{2\lambda}{\Lambda^2} - \delta) \right] \end{aligned}$$

$$\text{Pour } n \geq n_0 \quad (1-\alpha + (\frac{2\lambda}{\Lambda^2} - \delta)(1 + \frac{2\lambda}{\Lambda} - \delta\Lambda) K_2 c_n) \leq 1 - \beta < 1$$

$$\text{et } \|X_{n+1} - x\|^2 \leq (1 - \beta) \|X_n - x\|^2 + K_0 c_n.$$

En itérant cette inégalité à partir de n_0

$$\|X_{n_0+k} - x\|^2 \leq (1-\beta)^k \|X_{n_0} - x\|^2 + K_0 \left[c_{n_0} (1-\beta)^{k-1} + c_{n_0+1} (1-\beta)^{k-2} + \dots + c_{n_0+k-1} \right]$$

Le 2ème membre $\xrightarrow[k \rightarrow \infty]{} 0$ car c'est le terme général d'une série convergente.

Donc $X_n \xrightarrow[n \rightarrow \infty]{} x$ dans Φ fort.

IV.3.2.- Algorithme stochastique

IV.3.2.1.- Algorithme, hypothèses

On conserve les hypothèses déterministes du § IV.3.1 ($\underline{H_1}$, $\underline{H'_2}$, $\underline{H_3}$, $\underline{H_4}$) .

Les hypothèses stochastiques sont la traduction dans le cas présent des hypothèses faites au § II.3.

$$\underline{H'_5} : E[Z(x)] = 0$$

$$\underline{H'_{16}} : E[(\Delta Z(x,c), Px - P(x - \rho\Delta(x,c)))_{\Phi}] \leq$$

$$B E[(\Delta f(x,c), Px - P(x - \rho\Delta(x,c)))_{\Phi}]$$

où $\Delta(x,c)$, $\Delta f(x,c)$, $\Delta Z(x,c)$ ont même signification qu'au § IV.2.1.

B vérifie $\underline{H_7}$ ($0 \leq B \leq 1$) ou selon le cas $\underline{H'_7} : 0 \leq B \leq \frac{\lambda}{2\Lambda - \lambda} \leq 1$.

Remarque : Dans le cas d'un problème sans contrainte, il suffit pour vérifier H'_{16} que $E[(Z(x)-Z(y))^2] < B[f(x)-f(y)]^2$.

Algorithme

$$\left| \begin{array}{l} X_0 \in L^2(\underline{\omega}, \Phi) \quad (\text{par exemple un élément fixe de } K) \\ X_{n+1} = P(X_n - \rho_n \Delta_n) \quad \delta \leq \rho_n \leq \frac{2}{\Lambda} - \delta, \quad \delta \geq \frac{2}{\Lambda} \frac{1}{1+B} \end{array} \right.$$

Δ_n est défini comme au § IV.2.1 et les remarques 1 et 2 faites dans ce § demeurent valables.

IV.3.2.2.- Convergence de l'algorithme

On modifie comme au § IV.2.2 l'écriture de Δ_n

$$\Delta_n = \nabla f(X_n) + L_n + \sum_{i=1}^p \vec{u}_i \frac{Z(X_n + c_n \vec{u}_i) - Z(X_n - c_n \vec{u}_i)}{2c_n}$$

$$\underline{H'_2} \implies \|L_n\|_{\Phi} \leq c_n K_2.$$

Théorème 4.3.2.1.-

$$\begin{array}{l} \text{Si } \underline{H_1}, \underline{H'_2}, \underline{H_3}, \underline{H_4}, \underline{H'_5}, \underline{H'_{16}}, \underline{H_7} \text{ sont vérifiées,} \\ E[f(X_n)] \xrightarrow{n \rightarrow \infty} L = \inf_{x \in K} f(x) \end{array}$$

Etablissons d'abord le

Lemme 4.3.2.1.-

$$E[f(X_n)] \xrightarrow{n \rightarrow \infty} L'$$

Démonstration :

$$f(X_n) - f(X_{n+1}) = -(\nabla f(X_n), X_{n+1} - X_n)_{\Phi} - \frac{v'_n}{2} \|X_{n+1} - X_n\|_{\Phi}^2 \quad \lambda \leq v'_n \leq \Lambda.$$

D'où

$$(4.3.4) \quad E[f(X_n)] - E[f(X_{n+1})] = -(\nabla f(X_n), X_{n+1} - X_n)_{L^2} - \frac{v_n}{2} \|X_{n+1} - X_n\|_{L^2}^2 \quad \lambda \leq v_n \leq \Lambda$$

$$-\rho_n (\nabla f(X_n) + L_n + \sum_{i=1}^p \vec{u}_i \frac{Z(X_n + c_n \vec{u}_i) - Z(X_n - c_n \vec{u}_i)}{2c_n}, X_{n+1} - X_n)_{\Phi} = \mu'(\rho_n) \|X_{n+1} - X_n\|_{\Phi}^2$$

où $\mu'(\rho_n) \geq 1$ d'après (1.5.2)

$$E [\mu'(\rho_n) \|X_{n+1} - X_n\|_{\Phi}^2] = \mu''(\rho_n) \|X_{n+1} - X_n\|_{L^2}^2 \quad \text{où } \mu''(\rho_n) \geq 1$$

$$\underline{H'_{16}} \text{ donne : } \rho_n \left(\sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_n - X_{n+1} \right)_{L^2} \leq \rho_n B(\nabla f(X_n) + L_n, X_n - X_{n+1})_{L^2}.$$

Finalement

$$(4.3.5) \quad -\rho_n(1+B)(\nabla f(X_n) + L_n, X_{n+1} - X_n)_{L^2} = \mu(\rho_n) \|X_{n+1} - X_n\|_{L^2}^2$$

$$\text{où } \mu(\rho_n) \geq \mu''(\rho_n) \geq 1$$

$$\text{ou} \quad -(\nabla f(X_n), X_{n+1} - X_n)_{L^2} = \frac{\mu(\rho_n)}{\rho_n(1+B)} \|X_{n+1} - X_n\|_{L^2}^2 + (L_n, X_{n+1} - X_n)_{L^2}$$

(4.3.4) devient

$$(4.3.6) \quad E[f(X_n)] - E[f(X_{n+1})] = \left[\frac{\mu(\rho_n)}{\rho_n(1+B)} - \frac{v_n}{2} \right] \|X_{n+1} - X_n\|_{L^2}^2 + (L_n, X_{n+1} - X_n)_{L^2}$$

$(L_n, X_{n+1} - X_n)_{L^2}$ est le terme général d'une série absolument convergente.

En effet $|(L_n, X_{n+1} - X_n)_{L^2}| \leq \|L_n\|_{L^2} \|X_{n+1} - X_n\|_{L^2} \leq c_n K_2 \|X_{n+1} - X_n\|_{L^2}$ et

$\|X_{n+1} - X_n\|_{L^2}$ est borné car

$$\|X_{n+1} - X_n\|_{L^2}^2 \leq -\rho_n(1+B)(\nabla f(X_n) + L_n, X_{n+1} - X_n)_{L^2} \leq \rho_n(1+B) \|\nabla f(X_n) + L_n\|_{L^2} \|X_{n+1} - X_n\|_{L^2}$$

(4.3.5)

$$\|X_{n+1} - X_n\|_{L^2} \leq \rho_n(1+B) \|\nabla f(X_n) + L_n\|_{L^2} \quad \text{borné d'après } \underline{H'_2}$$

$$\frac{\mu(\rho_n)}{\rho_n(1+B)} - \frac{v_n}{2} \geq \frac{1}{(\frac{2}{\Lambda} - \delta)(1+B)} - \frac{\Lambda}{2} \geq 0 \quad \text{car } \delta \geq \frac{2}{\Lambda} \frac{B}{1+B}$$

Donc (4.3.6) s'écrit

$$E[f(X_n)] - E[f(X_{n+1})] = v_n + w_n$$

où $v_n \geq 0$ et $\sum_n |w_n| < +\infty$. On peut donc appliquer la proposition 1.5.2.2.

Lemme 4.3.2.2.-

$\sum_n \mu'(\rho_n) \|X_{n+1} - X_n\|_{L^2}^2 < +\infty$,
 $(\nabla f(X_n), X_n - X_{n+1})_{L^2}$ est le terme général
 d'une série convergente et $\{X_n\}$ est borné
 dans $L^2(\omega, \Phi)$

Démonstration :

Puisque $\sum_n v_n < +\infty$, $\sum_n \|X_{n+1} - X_n\|_{L^2}^2 < +\infty$

et $\sum_n \frac{v_n}{2} \|X_{n+1} - X_n\|_{L^2}^2 < +\infty$.

Comme $\frac{\mu(\rho_n)}{\rho_n(1+B)} \|X_{n+1} - X_n\|_{L^2}^2 = v_n + \frac{v_n}{2} \|X_{n+1} - X_n\|_{L^2}^2$,

on a : $\sum_n \mu(\rho_n) \|X_{n+1} - X_n\|_{L^2}^2 < +\infty$. Enfin $\mu(\rho_n) \geq \mu'(\rho_n) \geq 1$ et

$\sum_n |(L_n, X_{n+1} - X_n)_{L^2}| < +\infty$ implique avec (4.3.5) les derniers résultats.

Démonstration du théorème 4.3.2.1 :

Comme dans la démonstration du théorème 4.3.1.1 :

si $L' > L$, soit $z \in K$ tel que $f(z) < L'$ et posons $L' - f(z) = 2\varepsilon > 0$.

Soit $n \geq N(\varepsilon)$ tel que $-\varepsilon \leq E(f(X_n)) - L' \leq \varepsilon$

on a alors $0 > \varepsilon \delta \geq \rho_n [f(z) - E[f(X_n)]] \geq \rho_n (\nabla f(X_n), z - X_n)_{L^2}$

$0 > -\varepsilon \delta \geq (\rho_n \Delta_n, z - X_n)_{L^2} + (\rho_n L_n, X_n - z)_{L^2} + \rho_n \underbrace{\left(\sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n} \right)}_{=0}, X_n - z)_{L^2}$

$0 > -\varepsilon \delta \geq (-\rho_n \Delta_n, X_n - X_{n+1})_{L^2} + (\rho_n \Delta_n, z - X_{n+1})_{L^2} + (\rho_n L_n, X_n - z)_{L^2}$

$0 > -\varepsilon \delta \geq -\mu'(\rho_n) \|X_{n+1} - X_n\|_{L^2}^2 + (\rho_n L_n, X_n - z)_{L^2} + (X_n - X_{n+1}, z - X_{n+1})_{L^2}$
 $+ (X_{n+1} - (X_n - \rho_n \Delta_n), z - X_{n+1})_{L^2}$.

Le dernier terme est ≥ 0 d'après (1.5.1) . Donc :

$0 > -\varepsilon \delta \geq -\mu'(\rho_n) \|X_{n+1} - X_n\|_{L^2}^2 + (\rho_n L_n, X_n - z)_{L^2} + (X_n - X_{n+1}, z - X_{n+1})_{L^2}$.

D'après le lemme 4.3.2.2, le 2ème membre $\xrightarrow{n \rightarrow \infty} 0$ ce qui est absurde. D'où $L' = L$.

Théorème 4.3.2.2.-

Si, outre les hypothèses $\underline{H_1}, \underline{H'_2}, \underline{H_3}, \underline{H_4}, \underline{H'_5}, \underline{H'_{16}}, \underline{H'_7}$
 remplaçant $\underline{H_7}$, $\lambda > 0$ et $\delta < \rho_n < \frac{2\lambda}{\Lambda^2} - \delta$

alors $\{z \in K \mid f(z) = L\}$ est réduit à un point,
 soit x

et $X_n \xrightarrow{n \rightarrow \infty} x$ en moyenne quadratique

H₇ assure la compatibilité de $\frac{2}{\Lambda} \frac{B}{1+B} \leq \delta \leq \rho_n \leq \frac{2\lambda}{\Lambda^2} - \delta \leq \frac{2}{\Lambda} \frac{1}{1+B}$.

Comme dans les § précédents, les hypothèses faites assurent l'existence et l'unicité de x .

Reste à montrer : $X_n \xrightarrow[n \rightarrow \infty]{} x$ dans L^2 .

$$X_{n+1} = P(X_n - \rho_n (\nabla f(X_n) + L_n + \sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}))$$

$$x = P(x - \rho_n \nabla f(x))$$

Utilisons la proposition 1.5.1.2 avec

$$a = X_n - \rho_n (\nabla f(X_n) + L_n)$$

$$u = -\rho_n \sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}$$

$$b = x - \rho_n \nabla f(x)$$

$$\|X_{n+1} - x\|_{\Phi}^2 \leq \|X_n - x - \rho_n (\nabla f(X_n) + L_n - \nabla f(x))\|_{\Phi}^2 - 2\rho_n \left(\sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1} - x \right)_{\Phi}$$

Les transformations faites sur (4.3.3) sont applicables au 1er terme du 2ème membre. D'où :

$$(4.3.7) \quad \|X_{n+1} - x\|_{\Phi}^2 \leq (1-\beta) \|X_n - x\|_{\Phi}^2 + K_0 c_n - 2\rho_n \left(\sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_{n+1} - x \right)_{\Phi}$$

$$\text{ou} \quad \|X_{n+1} - x\|_{L^2}^2 \leq (1-\beta) \|X_n - x\|_{L^2}^2 + K_0 c_n + 2\rho_n \left(\sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, x - X_{n+1} \right)_{L^2}$$

$$\begin{aligned} \rho_n \left(\sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, x - X_{n+1} \right)_{L^2} &= \rho_n \underbrace{\left(\sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, x - X_n \right)_{L^2}}_0 + \rho_n \left(\sum_{i=1}^P \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_n - X_{n+1} \right)_{L^2} \\ &\leq \rho_n B \left| \left(\nabla f(X_n), X_n - X_{n+1} \right)_{L^2} \right| \end{aligned}$$

$$\text{Donc (4.3.8)} \quad \|X_{n+1} - x\|_{L^2}^2 \leq (1-\beta) \|X_n - x\|_{L^2}^2 + d_n$$

où $d_n > 0$ est tel que $\sum_n d_n < +\infty$ d'après le lemme 4.3.2.2.

En itérant (4.3.8) à partir de n_0 , on peut alors conclure comme au théorème 4.3.1.2.

Théorème 4.3.2.3.-

Sous les hypothèses du théorème 4.3.2.2,

$$X_n \xrightarrow[n \rightarrow \infty]{} x \text{ presque sûrement}$$

Démonstration :

$$\rho_n (\nabla f(X_n) + L_n + \Sigma_i^+ \frac{Z(+)-Z(-)}{2c_n}, X_n - X_{n+1})_\Phi = \mu(\rho_n) \|X_{n+1} - X_n\|_\Phi^2 \geq 0$$

Donc :

$$\begin{aligned} \rho_n (1+B) E \left[(\nabla f(X_n) + L_n, X_n - X_{n+1})_\Phi | X_1 \dots X_n \right] &\geq \\ \rho_n E \left[(\nabla f(X_n) + L_n + \Sigma_i^+ \frac{Z(+)-Z(-)}{2c_n}, X_n - X_{n+1})_\Phi | X_1 \dots X_n \right] &\geq 0 \quad (1) \end{aligned}$$

Modifions (4.3.7)

$$\begin{aligned} \|X_{n+1} - x\|_\Phi^2 &\leq (1-\beta) \|X_n - x\|_\Phi^2 + K_0 c_n + 2\rho_n (\Sigma_i^+ \frac{Z(+)-Z(-)}{2c_n}, x - X_n)_\Phi \\ &\quad + 2\rho_n (\nabla f(X_n) + L_n + \Sigma_i^+ \frac{Z(+)-Z(-)}{2c_n}, X_n - X_{n+1})_\Phi - 2\rho_n (\nabla f(X_n) + L_n, X_n - X_{n+1})_\Phi \end{aligned}$$

En prenant l'espérance conditionnelle, le 3ème terme du 2ème membre est nul et le dernier est négatif.

$$\begin{aligned} E \left[\|X_{n+1} - x\|_\Phi^2 | X_1 \dots X_n \right] &\leq \|X_n - x\|_\Phi^2 + K_0 c_n + 2 E \left[\mu(\rho_n) \|X_{n+1} - X_n\|_\Phi^2 | X_1 \dots X_n \right] \\ \text{ou } E \left[(\|X_{n+1} - x\|_\Phi^2 - K_0 \sum_{k=1}^n c_k - 2 \sum_{k=1}^n \mu(\rho_k) \|X_{k+1} - X_k\|_\Phi^2) | X_1 \dots X_n \right] \\ &\leq \|X_n - x\|_\Phi^2 - K_0 \sum_{k=1}^{n-1} c_k - 2 \sum_{k=1}^{n-1} \mu(\rho_k) \|X_{k+1} - X_k\|_\Phi^2 = -U_n \end{aligned}$$

Comme au théorème 2.3.2.3, $E^{a_n(U_{n+1})} \geq U_n$: on a une sous-martingale par rapport à $\{a_n\}$

$$\begin{aligned} |U_n| &\leq \|X_n - x\|_\Phi^2 + K_0 \sum_{k=1}^{n-1} c_k + 2 \sum_{k=1}^{n-1} \mu(\rho_k) \|X_{k+1} - X_k\|_\Phi^2 \\ E(|U_n|) &\leq \|X_n - x\|_{L^2}^2 + K_0 \sum_{k=1}^{n-1} c_k + 2 \sum_{k=1}^{n-1} \mu'(\rho_k) \|X_{k+1} - X_k\|_{L^2}^2 \text{ borné} \end{aligned}$$

(théorème 4.3.2.2 et lemme 4.3.2.2)

Donc il existe U tel que $U_n \xrightarrow[n \rightarrow \infty]{} U$ presque sûrement.

Posons $V = -U + K_0 \sum_{k=1}^{\infty} c_k$

$$\begin{aligned} \|X_n - x\|_\Phi^2 - 2 \sum_{k=1}^{n-1} \mu(\rho_k) \|X_{k+1} - X_k\|_\Phi^2 - K_0 \sum_{k=1}^{n-1} c_k &\xrightarrow[n \rightarrow \infty]{} -U \text{ p.s.} \\ -K_0 \sum_{k=n}^{\infty} c_k &\xrightarrow[n \rightarrow \infty]{} 0 \end{aligned}$$

(1) La remarque faite au théorème 4.2.2.3 demeure valable.

$$D'où \quad \|X_n - x\|_{\Phi}^2 - 2 \sum_{k=1}^{n-1} \mu(\rho_k) \|X_{k+1} - X_k\|_{\Phi}^2 \xrightarrow[n \rightarrow \infty]{} V \quad \text{p.s.}$$

On peut achever maintenant la démonstration comme au théorème 2.3.2.3.

IV.3.2.3.- Affaiblissement des hypothèses sur le bruit

De même qu'au chapitre II, les théorèmes 4.3.2.1, 4.3.2.2, 4.3.2.3 représentent une meilleure approche de la réalité que les théorèmes déterministes 4.3.1.1 et 4.3.1.2.

Cependant, ici aussi, l'hypothèse $\underline{H'_{16}}$ a peu de chances d'être vérifiée dans la pratique.

Il est plus réaliste de remplacer $\underline{H'_{16}}$ par

$$\underline{H_8} \quad \left\| \sum_{i=1}^p \vec{u}_i \frac{Z(x + c\vec{u}_i) - Z(x - c\vec{u}_i)}{2c} \right\|_{L^2}^2 \leq B^2 \quad \text{pour tout } x \in K \text{ et } c \text{ assez petit.}$$

Théorème 4.3.2.4.-

Si $\underline{H_1}$, $\underline{H'_2}$, $\underline{H_3}$, $\underline{H_4}$, $\underline{H'_5}$, $\underline{H_8}$ sont vérifiées, si en outre $\lambda > 0$ et $\exists N_1$ tel que $n \geq N_1 \implies \rho_n \leq \frac{\lambda}{\Lambda^2}$, alors $\limsup_n E(\|X_n - x\|_{\Phi}^2) \leq \frac{B^2}{\Lambda} (1 + K_3 \sum_{k=1}^{\infty} c_k)$ où $K_3 = K_2 \frac{\lambda}{\Lambda} (1 + \frac{\lambda}{\Lambda^2})$, K_2 étant défini au § IV.2.2

Démonstration :

Les hypothèses faites assurent l'existence et l'unicité de $x \in K$ tel que $f(x) = L$.

Utilisons la proposition 1.5.1.2 en choisissant a , b et u comme dans la démonstration du théorème 4.3.2.2.

$$\begin{aligned} \|X_{n+1} - x\|_{\Phi}^2 &\leq \|X_n - x - \rho_n (\nabla f(X_n) + L_n - \nabla f(x))\|_{\Phi}^2 - 2\rho_n (\sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n}, X_n - x - \rho_n (\nabla f(X_n) + L_n - \nabla f(x)))_{\Phi} \\ &\quad + \rho_n^2 \left\| \sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n} \right\|^2. \end{aligned}$$

Pour $n \geq N_1$, puisqu'alors $\rho_n \leq \frac{\lambda}{\Lambda^2}$,

$$1^{\text{er}} \text{ terme} \leq \left[(1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) + K_3 c_n \right] \|X_n - x\|_{\Phi}^2 + K_0 c_n.$$

En prenant l'espérance conditionnelle par rapport à X_n , le 2ème terme s'annule et $\rho_n^2 E \left[\left\| \sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n} \right\|^2 \middle| X_n \right] \leq \rho_n^2 B^2$ d'après H7.

Finalement :

$$(4.3.9) \quad \|X_{n+1} - x\|_{L^2}^2 \leq \left[(1 - 2\rho_n \lambda + \rho_n^2 \Lambda^2) + K_3 c_n \right] \|X_n - x\|_{L^2}^2 + K_0 c_n + \rho_n^2 B^2.$$

. Pour $n \geq N_1$, $\rho_n \leq \frac{\lambda}{\Lambda^2}$

Pour $n \geq N_2$, $K_3 c_n \leq \delta \lambda$ ($\leq \rho_n \lambda$) car $c_n \xrightarrow{n \rightarrow \infty} 0$

Pour $n \geq N_3$, $K_3 c_n \leq \frac{\lambda^2}{2\Lambda^2}$ ($\implies \lambda^2 - 2\Lambda^2 K_3 c_n \geq 0$)

Pour $n \geq N_4$ ($\geq N_3$) $\frac{\lambda - \sqrt{\lambda^2 - 2\Lambda^2 K_3 c_n}}{\Lambda^2} < \delta$ c'est possible

car l'expression $\xrightarrow{n \rightarrow \infty} 0$.

Soit $N = \sup (N_1, N_2, N_3, N_4)$.

Soit $\varepsilon > 0$ et $n_1(\varepsilon) \geq N$ tel que $c_k \leq \frac{1}{2} \frac{\varepsilon \delta \lambda}{K_0 + \varepsilon K_3}$ pour $k \geq n_1(\varepsilon)$.

Choisissons $n_2(\varepsilon) > n_1(\varepsilon)$ tel que

$$\frac{\varepsilon \Lambda^2 \delta}{2} \sum_{k=n_1(\varepsilon)}^{n_2(\varepsilon)-1} \rho_k \geq \|X_{n_1(\varepsilon)} - x\|_{L^2}^2 - \frac{B^2}{\Lambda^2} \left(1 + K_3 \sum_{k=1}^{n_1(\varepsilon)} c_k \right) .$$

. S'il existe $k \geq n_1(\varepsilon)$ tel que $\|X_k - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} \left(1 + K_3 \sum_{\ell=1}^{k-1} c_\ell \right) + \varepsilon$,

montrons que l'inégalité est vraie pour tout indice supérieur.

Pour un tel k , (4.3.9) donne

$$\begin{aligned} \|X_{k+1} - x\|_{L^2}^2 &\leq \left[(1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2) + K_3 c_k \right] \left[\frac{B^2}{\Lambda^2} \left(1 + K_3 \sum_{\ell=1}^{k-1} c_\ell \right) + \varepsilon \right] + K_0 c_k + \rho_k^2 B^2 \\ &\leq \frac{B^2}{\Lambda^2} \left[\left[1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2 + K_3 c_k \right] \left(1 + K_3 \sum_{\ell=1}^{k-1} c_\ell \right) + \rho_k^2 \Lambda^2 \right] \\ &\quad + \varepsilon \left[1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2 + K_3 c_k \right] + \varepsilon (\rho_k \lambda - K_3 c_k) \text{ car } k \geq n_1(\varepsilon) . \end{aligned}$$

Puisque $k \geq N_1$ et $k \geq N_2$, $\frac{B^2}{\Lambda^2} [] \leq \frac{B^2}{\Lambda^2} \left[1 + K_3 \sum_{\ell=1}^k c_\ell \right]$.

Puisque $k \geq N_1$,

$$\varepsilon \left[1 - 2\rho_k \lambda + \rho_k^2 \Lambda^2 + \rho_k \lambda \right] = \varepsilon \left[1 - (\rho_k \lambda - \rho_k^2 \Lambda^2) \right] \leq \varepsilon.$$

Finalement :

$$\|X_{k+1} - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} (1 + K_3 \sum_{\ell=1}^k c_\ell + \varepsilon).$$

. Si pour tout k : $n_1(\varepsilon) \leq k \leq n_2(\varepsilon) - 1$, $\|X_k - x\|_{L^2}^2 > \frac{B^2}{\Lambda^2} (1 + K_3 \sum_{\ell=1}^{k-1} c_\ell) + \varepsilon$,

on va majorer $\|X_{n_2(\varepsilon)} - x\|_{L^2}^2$

$$\|X_{k+1} - x\|_{L^2}^2 \leq \|X_k - x\|_{L^2}^2 + (-2\lambda\rho_k + \rho_k^2 \Lambda^2 + K_3 c_k) \|X_k - x\|_{L^2}^2 + K_0 c_k + \rho_k^2 B^2$$

$$(-2\lambda\rho_k + \rho_k^2 \Lambda^2 + K_3 c_k) \leq -\rho_k \lambda + \rho_k^2 \Lambda^2 \leq 0$$

$$(k \geq N_2) \quad (k \geq N_1)$$

$$\|X_{k+1} - x\|_{L^2}^2 \leq \|X_k - x\|_{L^2}^2 + (-2\lambda\rho_k + \rho_k^2 \Lambda^2 + K_3 c_k) \left[\frac{B^2}{\Lambda^2} (1 + K_3 \sum_{\ell=1}^{k-1} c_\ell) + \varepsilon \right] + K_0 c_k + \rho_k^2 B^2$$

$$\leq \|X_k - x\|_{L^2}^2 + \frac{B^2}{\Lambda^2} \left[(1 + K_3 \sum_{\ell=1}^{k-1} c_\ell) (-2\lambda\rho_k + \rho_k^2 \Lambda^2 + K_3 c_k) + \rho_k^2 \Lambda^2 \right]$$

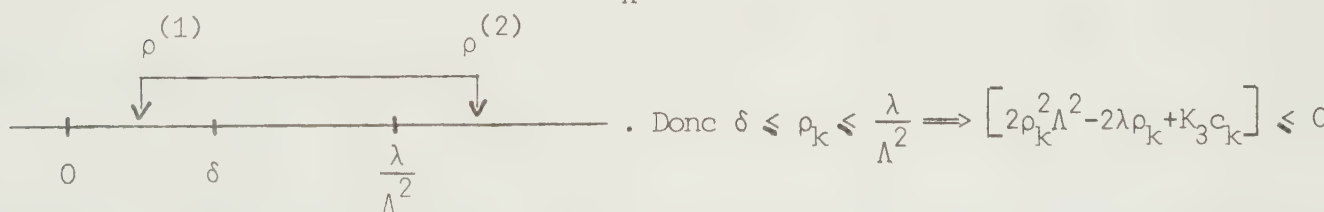
$$+ \varepsilon (-2\lambda\rho_k + \rho_k^2 \Lambda^2 + K_3 c_k) + \varepsilon \left(\frac{\rho_k \lambda}{2} - K_3 c_k \right)$$

$$\left[\right] \leq (1 + K_3 \sum_{\ell=1}^{k-1} c_\ell) \left[+ 2\rho_k^2 \Lambda^2 - 2\lambda\rho_k + K_3 c_k \right].$$

Puisque $k \geq N_3$, les racines du trinôme en ρ_k sont : $\rho^{(1)} = \frac{\lambda - \sqrt{\lambda^2 - 2\Lambda^2 K_3 c_k}}{\Lambda^2}$

$$\rho^{(2)} = \frac{\lambda + \sqrt{\lambda^2 - 2\Lambda^2 K_3 c_k}}{\Lambda^2}$$

$k \geq N_4 \implies \rho^{(1)} \leq \delta$. De plus $\rho^{(2)} \geq \frac{\lambda}{\Lambda^2}$



$$\|X_{k+1} - x\|_{L^2}^2 \leq \|X_k - x\|_{L^2}^2 + \varepsilon \left(-\frac{3}{2}\lambda\rho_k + \rho_k^2 \Lambda^2 \right) \leq \|X_k - x\|_{L^2}^2 - \varepsilon \rho_k^2 \frac{\Lambda^2}{2} \leq \|X_k - x\|_{L^2}^2 - \frac{\varepsilon \Lambda^2 \delta}{2} \rho_k$$

($k \geq N_1$)

En itérant entre $n_1(\varepsilon)$ et $n_2(\varepsilon) - 1$

$$\begin{aligned} \|X_{n_2(\varepsilon)} - x\|_{L^2}^2 &\leq \|X_{n_1(\varepsilon)} - x\|_{L^2}^2 - \frac{\varepsilon \Lambda^2 \delta}{2} \sum_{k=n_1(\varepsilon)}^{n_2(\varepsilon)-1} \rho_k \leq \\ &\leq \|X_{n_1(\varepsilon)} - x\|_{L^2}^2 - \|X_{n_1(\varepsilon)} - x\|_{L^2}^2 + \frac{B^2}{\Lambda^2} (1 + K_3 \sum_{k=1}^{n_1(\varepsilon)} c_k) \\ &\text{(choix de } n_2(\varepsilon)). \end{aligned}$$

$$\text{Donc } \|X_{n_2(\varepsilon)} - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} (1 + K_3 \sum_{k=1}^{n_2(\varepsilon)-1} c_k) \quad \text{car } n_2(\varepsilon) - 1 \geq n_1(\varepsilon)$$

$$\text{Pour } n \text{ assez grand on a donc toujours } \|X_n - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} (1 + K_3 \sum_{k=1}^{\infty} c_k) + \varepsilon.$$

$$\text{D'où } \limsup_n \|X_n - x\|_{L^2}^2 \leq \frac{B^2}{\Lambda^2} (1 + K_3 \sum_{k=1}^{\infty} c_k).$$

Donnons à présent un corollaire du théorème 4.3.2.4 utile pour certaines applications où $\underline{H}_8$ ne serait pas vérifiée. Ce corollaire relève du même principe que le corollaire 2.3.2.5 du théorème 2.3.2.4.

Corollaire 4.3.2.5.-

Sous les mêmes hypothèses qu'au théorème 4.3.2.4 sauf à remplacer $\underline{H}_8$ par :

$$\left\| \sum_{i=1}^p \vec{u}_i \frac{Z(x + c\vec{u}_i) - Z(x - c\vec{u}_i)}{2c} \right\|_{L^2}^2 \leq d_1 + d_2 \|x\|^2$$

pour c assez petit et $\rho_n \leq \frac{\lambda}{\Lambda^2 + 2d_2}$, alors

$$\limsup_n \|X_n - x\|_{L^2}^2 \leq \frac{d_1 + d_2 \|x\|^2}{\Lambda^2 + 2d_2} (1 + K_3 \sum_{k=1}^{\infty} c_k)$$

Démonstration :

Les hypothèses faites assurent que

$$\rho_n^2 \left\| \sum_{i=1}^p \vec{u}_i \frac{Z(+)-Z(-)}{2c_n} \right\|_{L^2}^2 \leq \rho_n^2 (d_1 + d_2 \|X_n\|_{L^2}^2) \leq 2\rho_n^2 d_2 \|X_n - x\|_{L^2}^2 + \rho_n^2 (d_1 + 2d_2 \|x\|^2)$$

On peut donc poursuivre la démonstration comme au théorème 4.3.2.4 en remplaçant Λ^2 par $\Lambda^2 + 2d_2$ et B^2 par $d_1 + 2d_2 \|x\|^2$ et en prenant garde que K_3 lui, n'est pas modifié.

Détaillons un exemple proposé par M. Engelhardt [1] où le corollaire 4.3.2.5 est applicable.

Soit un problème de stock où la demande journalière d'un produit est T de fonction de répartition $p(T \leq t) = F(t)$.

Un jour donné, le vendeur stocke $x > 0$. Le prix de vente unitaire est S .

Les coûts de vente sont de trois types :

- Le prix d'achat de la marchandise : c_0 par unité
- la gestion du point de vente : c_1 fixe
- le coût $\chi(w)$ de l'excédent de stock w car la denrée est périssable.

Le bénéfice sera

$$Y(x) = S \min(x, T) - c_0 x - c_1 - \chi(x-T)$$

et on veut optimiser

$$f(x) = E(Y(x)) = S \left[x \left(1 - \frac{c_0}{S} \right) - \frac{c_1}{S} - \int_0^x F(t) dt \right] - E(\chi(x-T)) .$$

Pour préciser plus les données, on choisit

$$\begin{aligned} F(t) &= 0 & t < t_0 \\ &= 1 - \left(\frac{2t}{t_0} \right)^{-\beta} & t \geq t_0 > 0 \quad 0 < \beta < 1 \end{aligned}$$

$$\begin{aligned} \chi(w) &= 0 & 0 \leq w < w_0 \\ &= c_2 w & w \geq w_0 > 0, \quad c_2 > 0 . \end{aligned}$$

Dans ces conditions χ est discontinue en w_0

et $E(\chi(x-T))$ est discontinue en $w_0 + t_0$ qui réalise justement le maximum de $f(x)$.

En posant $Z(x) = Y(x) - E(Y(x))$, il est clair que $E(Z(x)^2)$ n'est pas borné, mais est une fonction croissante de x .

L'hypothèse H_0 n'est pas vérifiée, mais le corollaire 4.3.2.5 est lui, applicable.

IV.4.- Comparaison avec les résultats antérieurs

Comme cela a été dit au § II.4, les résultats publiés jusqu'ici par d'autres auteurs, se situent à notre connaissance, dans le cadre : - problème sans contrainte
- Φ de dimension finie

Le calcul d'une valeur approchée du gradient par différences nous a aussi imposé la condition : Φ de dimension finie. La supériorité des méthodes traitées aux § IV.2 et IV.3 est donc seulement la possibilité de traiter des problèmes avec contraintes.

Dans le contexte présent, les travaux publiés antérieurement concernent tous (à notre connaissance) l'algorithme de Kiefer Wolfowitz [1] qui est exactement celui étudié au § IV.2 dans le cas d'un problème sans contrainte ($P = 1$ 'identité).

Les conditions sur $\{\rho_n\}$ et $\{c_n\}$ sont celles du § IV.2.

Comme au § II.4, les auteurs utilisent selon les cas deux techniques pour obtenir la convergence sûre :

- la théorie des martingales
- le lemme de Borel-Cantelli

Les théorèmes de convergence presque sûre obtenue à l'aide de la théorie des martingales donnent lieu à des conditions très techniques.

Nous n'en citons pas, mais on pourra consulter par exemple J. Sacks [1].

L'utilisation du lemme de Borel-Cantelli conduit au

Théorème 4.4.1.- (J.H. Venter)

Si $\underline{L_1}$: f est bornée inférieurement sur Φ
 $\underline{L_2}$: f admet des dérivées bornées jusqu'au 2ème ordre sur Φ
 $\underline{L_3}$: θ est un minimum local de f
 $\underline{L_4}$: $\forall x \in \Phi, x \neq \theta, \nabla f(x) \neq 0$
 $\underline{L'_5}$: $E[(Z(x))^2] < K_0 \quad \forall x \in \Phi$
 alors presque sûrement, $X_n \xrightarrow[n \rightarrow \infty]{} \theta$ ou $\|X_n\|_{\Phi} \xrightarrow[n \rightarrow \infty]{} +\infty$
 Si de plus :
 $\underline{L_6}$: $\exists T > 0$ tel que $\|x - \theta\|_{\Phi} > T \implies (x - \theta, \nabla f(x))_{\Phi} \geq 0$
 alors $X_n \xrightarrow[n \rightarrow \infty]{} \theta$ presque sûrement.

Pour la démonstration, nous renvoyons à J.H. Venter [1] .

A l'exception de $\underline{L'_5}$ évidemment différente de $\underline{L_5}$, les hypothèses du théorème 4.4.1 sont identiques à celles du théorème 2.4.2.1.

Résumons la comparaison entre les théorèmes 4.2.2.2 et 4.4.1 dans un tableau :

Théorème 4.2.2.2		Théorème (4.4.1)
Domaine de validité	Problème avec contraintes Φ de dimension finie	Problème sans contrainte Φ de dimension finie
pas d'itération	$c_n, \rho_n > 0, \sum_n \rho_n = +\infty, \sum_n \rho_n c_n < +\infty$ $\sum_n \left(\frac{\rho_n}{c_n}\right)^2 < +\infty, c_n \xrightarrow{n \rightarrow \infty} 0, \rho_n \xrightarrow{n \rightarrow \infty} 0$	$\longrightarrow$ idem
comparaison des hypothèses	$\underline{H_1}$	$\Longleftrightarrow \underline{L_1}$
	$\underline{H'_2}$	$\Longleftrightarrow \underline{L_2}$
	$\underline{H_1} + \underline{H'_2} + \underline{H_3} + \underline{H_4}$	$\implies \underline{L_1} + \underline{L_2} + \underline{L_3} + \underline{L_4}$
	$\underline{H_{16}}$	$\Longleftrightarrow \underline{L'_5}$
	$\underline{H_4}$	$\implies \underline{L_6}$
vitesse de convergence	lente	$\longrightarrow$ idem
type de convergence	presque sûre et en moyenne quadratique	presque sûre

5

méthodes
stochastiques
de dualité

V.1.- Enoncé du problème, données

V.1.1.- Données, hypothèses

Nous nous situons dans un contexte théorique envisagé par J.L. Lions, R. Tremolières, R. Glowinski [1] (dans le cas quadratique) ou par A. Bensoussan, J.L. Lions, R. Temam [1].

Soit $\mathcal{U}$ un espace de Hilbert séparable sur $\mathbb{R}$ et $\mathcal{U}_{ad} \subset \mathcal{U}$, $\mathcal{U}_{ad}$ convexe fermé.

Soit $J : v \longmapsto J(v)$ une fonctionnelle vérifiant

$$\mathcal{U} \longrightarrow \mathbb{R}$$

$$J_1 : J \text{ est } C_2$$

$$J_2 : (\nabla J(u) - \nabla J(v), u-v)_{\mathcal{U}} \geq m \|u-v\|_{\mathcal{U}}^2, \quad m > 0$$

$$J_3 : \|\nabla J(u) - \nabla J(v)\|_{\mathcal{U}} < M \|u-v\|_{\mathcal{U}}.$$

Soient $\mathcal{U}_{ad}$ et $\mathcal{U}'_{ad}$ deux convexes fermés de $\mathcal{U}$.

On peut repérer $\mathcal{U}'_{ad}$ par la donnée

- de Λ , cône convexe de sommet 0 dans un espace vectoriel normé séparable Ψ à préciser

- et d'une fonction $(v, q) \longrightarrow \Phi(v, q)$

$$\mathcal{U} \times \Lambda \longrightarrow \mathbb{R}$$

convexe, s.c.i. en v et telle que $\Phi(v, \rho q) = \rho \Phi(v, q) \quad \forall \rho \geq 0$
concave, s.c.s. en p .

$$\text{Alors } v \in \mathcal{U}'_{ad} \iff \Phi(v, q) \leq 0 \quad \forall q \in \Lambda.$$

Pour simplifier et parce que ce sera suffisant les applications projetées, nous supposons que Ψ est un espace de Hilbert identifié à son dual et que $\Phi(v, q)$ est de la forme $\Phi(v, q) = (q, Bv)_{\Psi}$

où $B : v \longmapsto B(v)$ vérifie $\|B(v) - B(v')\|_{\Psi} \leq C \|v-v'\|_{\mathcal{U}} \quad \forall v, v' \in \mathcal{U}'_{ad}$.

$$\mathcal{U}'_{ad} \longrightarrow \Psi$$

Pour inclure dans la théorie les problèmes de minimisation de fonctionnelles non différentiables, nous supposons seulement que Λ est convexe fermé.

Enfin, nous ferons l'hypothèse de qualification :

$$Q : \exists v_0 \in \mathcal{U}'_{ad} \text{ tel que } (q, B(v_0))_{\Psi} < 0 \quad \forall q \in \Lambda.$$

V.1.2.- Problème

On considère le problème
$$\min_{v \in \mathcal{U}_{ad} \cap \mathcal{U}'_{ad}} J(v) .$$

L'introduction du repérage de $\mathcal{U}'_{ad}$ permet de reformuler ce problème :
problème primal :

$$(5.1.1) \quad \inf_{v \in \mathcal{U}_{ad}} \sup_{q \in \Lambda} \{J(v) + (q, B(v))_{\Psi}\}$$

auquel on associe le

problème dual

$$(5.1.2) \quad \sup_{q \in \Lambda} \inf_{v \in \mathcal{U}_{ad}} \{J(v) + (q, B(v))_{\Psi}\}$$

V.1.3.- Théorèmes d'existence et d'unicité, reformulation du problème dual

Posons $\mathcal{L}(v, q) = J(v) + (q, B(v))_{\Psi}$.

Remarques sur les hypothèses :

- 1- $m > 0$ n'est pas nécessaire pour un théorème d'existence. Il ne sera requis que pour l'unicité
- 2- $\underline{J}_2 \Rightarrow \|u\| \longrightarrow +\infty \Rightarrow J(u) \longrightarrow +\infty$
- 3- $\underline{J}_1$ sera nécessaire seulement pour utiliser le théorème des fonctions implicites. Pour l'instant, il suffit que J soit C^1 .
- 4- $\underline{Q} \Rightarrow \mathcal{U}'_{ad}$ admet au moins un point intérieur.
- 5- On peut remplacer $\underline{Q}$ par une autre hypothèse de qualification comme on le fait couramment en programmation mathématique.

Enonçons maintenant sans démonstration quelques résultats classiques. On pourra consulter J. Céa [1] ou J. Céa, R. Glowinski [1] .

Théorème 5.1.3.1.-

(5.1.3)

Sous les hypothèses $\underline{Q}$, $\underline{J}_1$, $\underline{J}_2$, $\underline{J}_3$, il existe
 $(u, p) \in \mathcal{U}_{ad} \times \Lambda$ solution de $\mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(v, p)$
 $\forall (v, q) \in \mathcal{U}_{ad} \times \Lambda$.

Corollaire 5.1.3.2.-

Un couple (u, p) solution de (5.1.3) est aussi solution de (5.1.4) et réciproquement

$$(5.1.4) \quad \begin{cases} u \in \mathcal{U}_{ad}, p \in \Lambda \\ (\nabla J(u), v-u)_{\mathcal{U}} + (p, B(v)-B(u))_{\Psi} \geq 0 \quad \forall v \in \mathcal{U}_{ad} \\ (q-p, B(u))_{\Psi} \leq 0 \quad \forall q \in \Lambda \end{cases}$$

Ce corollaire montre que la théorie qui va suivre s'applique aux problèmes posés en termes d'inéquations : Remplaçons dans (5.1.4) $\nabla J(u)$ par $A(u)$ où A vérifie les hypothèses :

A_1 : A est C^1

A_2 : $(A(u)-A(v), u-v)_{\mathcal{U}} \geq m \|u-v\|_{\mathcal{U}}^2 \quad (m > 0)$

A_3 : $\|A(u) - A(v)\|_{\mathcal{U}} \leq M \|u-v\|_{\mathcal{U}}$

Les paragraphes suivants donneront des moyens pour résoudre le problème :

$$(5.1.5) \quad \begin{cases} u \in \mathcal{U}_{ad}, p \in \Lambda \\ (A(u), v-u)_{\mathcal{U}} + (p, B(v) - B(u))_{\Psi} \geq 0 \quad \forall v \in \mathcal{U}_{ad} \\ (q-p, B(u))_{\Psi} \leq 0 \quad \forall q \in \Lambda \end{cases}$$

Remarque : on a implicitement identifié $\mathcal{U}$ à son dual. Ce n'est pas essentiel pour la suite. On pourrait écrire $\langle \nabla J(u), v-u \rangle_{\mathcal{U}^*, \mathcal{U}}$ ou $\langle A(u), v-u \rangle_{\mathcal{U}^*, \mathcal{U}}$.

Corollaire 5.1.3.3.-

Si (u_1, p_1) et (u_2, p_2) sont 2 solutions de (5.1.3), alors $u_1 = u_2$.

(On utilise ici $m > 0$)

Donc :

Corollaire 5.1.3.4.-

Si (u, p) est solution de (5.1.3) u est la solution unique du problème primal.

Corollaire 5.1.3.5.-

Si p est une solution du problème dual et si u minimise $J(v) + (p, B(v))_{\Psi}$ sur $\mathcal{U}_{ad}$, alors (u, p) vérifie (5.1.3) et u est la solution du problème primal.

On peut maintenant reformuler le problème dual

$$(5.1.6) \quad \left[\begin{array}{l} \sup_{q \in \Lambda} [J(u_q) + (q, B(u_q))_\Psi] \\ u_q \text{ défini par :} \\ q \in \Lambda, u_q \in \mathcal{U}_{ad} \\ (\forall J(u_q), v - u_q)_{\mathcal{U}} + (q, B(v) - B(u_q))_\Psi \geq 0 \quad \forall v \in \mathcal{U}_{ad} \end{array} \right]$$

V.2.- Exemples

On pourra trouver dans A. Bensoussan, J.L. Lions, R. Temam [1] des exemples économiques du problème envisagé.

Donnons ici des exemples en théorie du contrôle.

V.2.1.- Remarques générales (cf. J.L. Lions [1])

Soit le problème

$$(5.2.1) \quad \inf_{v \in \mathcal{U}_{ad}} J(v)$$

où $J(v) = \frac{1}{2} \pi(v, v) - L(v) + J(0)$ est une fonctionnelle quadratique vérifiant :

- $\pi(u, v)$ est une forme bilinéaire symétrique, continue sur $\mathcal{U}$ telle que $\pi(v, v) \geq m \|v\|^2$, $m > 0 \quad \forall v \in \mathcal{U}$.
- $L(v)$ est une forme linéaire continue sur $\mathcal{U}$.

Nous voulons remplacer le problème (5.2.1) par un problème de point-selle équivalent.

Introduisons la fonction $v \mapsto F(v)$ définie par

$$F(v) = \begin{cases} -L(v) & v \in \mathcal{U}_{ad} \\ +\infty & v \notin \mathcal{U}_{ad} \end{cases}$$

$$(5.2.1) \iff (5.2.2) \quad \inf_{v \in \mathcal{U}} \left[\frac{1}{2} \pi(v, v) + F(v) \right]$$

F est convexe, s.c.i., à valeurs dans $]-\infty, +\infty]$, $F \not\equiv +\infty$.

Soit $F^*(w)$ la fonction duale de F par rapport à π :

$$F^*(w) = \sup_{v \in \mathcal{U}} [\pi(w, v) - F(v)].$$

Dans ces conditions, on sait (J.L. Lions [1]) que

$$(5.2.3) \quad \left| \begin{array}{l} \inf_{v \in \mathcal{U}} \left[\frac{1}{2} \pi(v,v) + F(v) \right] + \inf_{v \in \mathcal{U}} \left[\frac{1}{2} \pi(v,v) + F^*(v) \right] = 0, \\ \text{Si } u \text{ est l'élément qui réalise la 1ère borne inférieure,} \\ \text{alors la 2ème borne inférieure est atteinte en } -u. \end{array} \right.$$

En posant $\mathcal{L}(v,w) = \frac{1}{2} \pi(v,v) + \pi(v,w) - F(w)$

(5.2.3) revient à résoudre le problème de point-selle :

$$\mathcal{L}(u,w) \leq \mathcal{L}(u,v) \leq \mathcal{L}(t,v) \quad \forall w \in \mathcal{U}, \quad \forall t \in \mathcal{U}.$$

Remarque : On peut associer au problème (5.2.1) d'autres problèmes de point-selle.

Par exemple, remarquons que $F(v) = +\infty$ si $v \notin \mathcal{U}_{ad}$.

$$\text{Donc } F^*(w) = \sup_{v \in \mathcal{U}_{ad}} (\pi(w,v) - F(v)) = \sup_{v \in \mathcal{U}_{ad}} (\pi(w,v) + L(v)).$$

En posant $\mathcal{L}'(v,w) = \frac{1}{2} \pi(v,v) + \pi(v,w) + L(w)$

(5.2.3) équivaut donc à résoudre le problème de point-selle

$$\mathcal{L}'(u,w) \leq \mathcal{L}'(u,v) \leq \mathcal{L}'(t,v) \quad \forall w \in \mathcal{U}_{ad}, \quad \forall t \in \mathcal{U}.$$

V.2.2.- Un problème de contrôle elliptique.

On reprend le problème posé au § I.1.1 que nous avons déjà étudié selon un autre point de vue au § II.5.1.

Le problème est donc : $\inf_{v \in \mathcal{U}_{ad}} J(v)$

où $J(v) = \| \mathcal{E}y(v) - z_d \|_{\mathcal{H}}^2 + (Nv, v)_{\mathcal{U}}$

$$J(v) \text{ s'écrit : } \frac{1}{2} J(v) = \frac{1}{2} \pi(v,v) - L(v) + \frac{1}{2} \| z_d - \mathcal{E}y(0) \|_{\mathcal{H}}^2$$

avec $\pi(u,v) = (\mathcal{E}(y(u) - y(0)), \mathcal{E}(y(v) - y(0)))_{\mathcal{H}} + (Nu, v)_{\mathcal{U}}$

$$L(v) = (z_d - \mathcal{E}y(0), \mathcal{E}(y(v) - y(0)))_{\mathcal{H}}$$

$$\pi(v,v) \geq v \|v\|^2 \quad (\text{propriété de } N)$$

$v \mapsto L(v)$ est linéaire car $v \mapsto y(v)$ est affine.

On est donc bien dans le contexte du § V.2.1 (cf. J.L. Lions [1]).

Soit $S \in \mathcal{L}(\mathcal{U}, V)$ défini par $Sv = y(v) - y(0)$.

Etudions sur cet exemple comment se transmet, dans le problème de point-selle associé, un bruit sur l'observation.

On peut écrire $\Pi(u,v) = (Mu,v)_{\mathcal{U}}$

$$L(v) = (L,v)_{\mathcal{U}}$$

où $T = S^* \mathcal{G}^* \mathcal{G} S + N$

et $L = S^* \mathcal{G}^*(z_d - \mathcal{G}y(0))$.

L'observation fournit la connaissance de $\mathcal{G}y(v) + \eta(v) = \mathcal{G}Sv + \eta(v)$.

On en déduit donc $\Pi Iv = L + Z(v)$

où $Z(v) = S^* \mathcal{G}^* \eta(v)$ est un bruit dont les propriétés stochastiques se déduisent aisément de celles de $\eta(v)$.

V.2.3.- Un problème de contrôle parabolique

Soit un problème parabolique avec (par exemple) contrôle et observation frontière. (cf. J.L. Lions [1]).

On reprend une partie du contexte du § I.1.2.

On considère le système :

$$\begin{cases} \frac{\partial y(v)}{\partial t} + Ay(v) = f & \text{dans } Q = \Omega \times [0,T] \\ \frac{\partial y(v)}{\partial \nu_A} = v & \text{sur } \Sigma = \partial \Omega \\ y(x,0;v) = y_0(x) & , \quad x \in \Omega \end{cases}$$

et le problème de contrôle

$$\inf_{v \in \mathcal{U}_{ad}} J(v) = \inf_{v \in \mathcal{U}_{ad}} \int_{\Sigma} (y(v) - z_d)^2 d\Sigma + (Nv,v)_{L^2(\Sigma)}.$$

L'opérateur $\mathcal{G}$ d'observation est donc ici l'opérateur trace sur Σ .

Pour se placer dans le contexte du § V.2.1, on peut poser :

$$\begin{aligned} \pi(u,v) &= \int_{\Sigma} (y(u) - y(0)) (y(v) - y(0)) d\Sigma + (Nu,v)_{L^2(\Sigma)} \\ L(v) &= \int_{\Sigma} (z_d - y(0)) (y(v) - y(0)) d\Sigma. \end{aligned}$$

Ici encore, l'observation bruitée $\mathcal{G}y(v) + \eta(v) = y(v)|_{\Sigma} + \eta(v)$ permet de déduire $\Pi Iv = L + Z(v)$.

Ces deux exemples montrent l'intérêt d'étudier des algorithmes de recherche de point-selle utilisant une observation entachée d'un bruit.

5.3.- Algorithmes stochastiques de recherche de points-selle, généralités.

Nous allons étudier dans la suite de ce chapitre deux types de méthodes : une méthode "de type Uzawa" et une méthode "de type Arrow, Hurwicz" (cf. K.J. Arrow, L. Hurwicz, H. Uzawa [1]).

V.3.1.- Algorithme d'Uzawa

Dans les algorithmes de type Uzawa, l'idée de base est la suivante : On résout le problème dual par une méthode de gradient et on récupère comme "sous-produit" une approximation de la solution optimale du problème primal.

L'algorithme déterministe est défini par :

$$\left| \begin{array}{l} p_0 \text{ quelconque } \in \Lambda \\ p_n \text{ étant connue, soit } u_n \text{ l'élément de } \mathcal{U}_{ad} \text{ qui minimise} \\ J(v) + (p_n, B(v))_{\Psi} \text{ sur } \mathcal{U}_{ad} \\ \text{puis } p_{n+1} = p_{\Lambda}(p_n + \rho_n B(u_n)) \\ p_{\Lambda} : \text{projecteur sur } \Lambda, \rho_n \text{ à choisir.} \end{array} \right.$$

On suppose que l'observation permet de calculer à q fixé le minimum u_q de $g_q(v) = J(v) + (q, B(v))_{\Psi}$ sur $\mathcal{U}_{ad}$ par exemple par un algorithme d'optimisation déterministe. Comme on n'obtiendra pas en général la solution exacte, on peut considérer qu'on trouve seulement $u_q + Z(q)$ où $Z(q)$ est une erreur que nous traiterons d'un point de vue stochastique. B. Martinet [1], étudie ce type d'erreur d'un point de vue déterministe. Voir aussi R.T. Rockafellar [1].

En outre l'algorithme de calcul de u_q peut être lui-même un algorithme d'optimisation stochastique si l'observation fournit seulement $\nabla g_q(v)$ entaché d'un bruit.

Nous supposons donc que l'observation permet de trouver (une réalisation de) $u_q + Z(q)$

où $Z(q)$ est une variable aléatoire vérifiant les hypothèses

$$\begin{array}{l} \underline{S_1} \quad E[B(u_q + Z(q))] = B(u_q) \\ \underline{S_2} \quad E[\|Z(q)\|^2] \leq B^2 \end{array} .$$

L'algorithme stochastique sera donc

$$\begin{aligned}
 (5.3.1) \quad & \left| \begin{array}{l}
 p_0 \text{ quelconque } \in \Lambda \\
 u_0 = u_{p_0} + Z(p_0) \\
 p_n \text{ étant connu, } U_n = u_{p_n} + Z_n \\
 p_{n+1} = P_\Lambda (p_n + \rho_n B(U_n)) \\
 \text{où } u_{p_n} \text{ est l'élément de } \mathcal{U}_{ad} \text{ qui minimise } J(v) + (p_n, B(v))_\Psi \\
 \text{sur } \mathcal{U}_{ad} \text{ et où } Z_n \text{ est une variable aléatoire qui à } p_n = q \\
 \text{fixé a même loi que } Z(q) .
 \end{array} \right.
 \end{aligned}$$

Explicitons cet algorithme sur l'exemple V.2.2.

Repérons $\mathcal{U}_{ad}$ à l'aide de $\Lambda \subset \Psi$:

$$v \in \mathcal{U}_{ad} \iff (q, B(v))_\Psi \leq 0 \quad \forall q \in \Lambda \subset \Psi .$$

$$\text{On pose } \mathcal{L}(v, q) = \| \mathcal{E}y(v) - z_d \|^2 + (Nv, v) + (q, B(v))_\Psi .$$

$$\text{Soit } p_n \in \Lambda .$$

On résout le problème :

$$\inf_{v \in \mathcal{U}} \mathcal{L}(v, p_n) \quad \text{à l'aide de l'observation entachée d'un bruit}$$

$$\mathcal{E}y(v) + \eta(v) .$$

Soit u_{p_n} la résolution de ce problème.

En raison de la présence de η , on pourra seulement trouver $U_n = u_{p_n} + Z_n$.

On pose ensuite

$$p_{n+1} = P_\Lambda (p_n + \rho_n B(U_n)) .$$

On retrouve sur cet exemple la situation habituelle dans la méthode d'Uzawa : L'algorithme n'est pas complètement déterminé. Le mode de calcul de U_n reste à choisir. La méthode d'Arrow Hurwicz est une méthode pour préciser ce point : on remplace le calcul de u_{p_n} par le calcul d'une valeur approchée à l'aide d'une seule itération de gradient.

V.3.2.- Algorithme d'Arrow Hurwicz

On suppose à présent que $\mathcal{U}_{ad} \cap \mathcal{U}'_{ad} = \mathcal{U}'_{ad}$ i.e. $\mathcal{U}_{ad} = \mathcal{U}$ et que $B \in \mathcal{L}(\mathcal{U}, \Psi)$

$g_q(v)$ admet donc le gradient :

$$\nabla g_q(v) = \nabla J(v) + B^* q \quad .$$

L'observation nous fournit pour tout v (une réalisation de) $\nabla J(v) + Z(v)$.
L'algorithme stochastique sera donc :

$$(5.3.2) \quad \left[\begin{array}{l} u_0 \in \mathcal{U} \text{ quelconque, } p_0 \in \Lambda \text{ quelconque} \\ U_{n+1} = U_n - \rho_n^{(1)} (\nabla J(U_n) + B^* P_n + Z_n) \\ P_{n+1} = P_\Lambda (P_n + \rho_n^{(2)} B^* U_{n+1}) \\ \text{où } Z_n \text{ est une variable aléatoire } \in L^2(\underline{\omega}, \mathcal{U}') \text{ dont la loi à} \\ U_n = v \text{ fixé est identique à celle de } Z(v), \rho_n^{(1)} \text{ et } \rho_n^{(2)} \text{ sont} \\ \text{à définir plus loin.} \end{array} \right.$$

Explicitons cet algorithme sur l'exemple V.2.2

$$\nabla J(v) = 2 \Pi v - 2L$$

$$\Pi = S^* \mathcal{E}^* \mathcal{E} S + N$$

$$L = S^* \mathcal{E}^* (z_d - \mathcal{E} y(0)) \quad .$$

L'observation nous fournit $\mathcal{E} y(v) + \eta(v) = \mathcal{E}[Sv + y(0)] + \eta(v)$.

$$\begin{aligned} \text{On en déduit } 2S^* \mathcal{E}^* [\mathcal{E} Sv + \eta(v)] + 2Nv - 2S^* \mathcal{E}^* [z_d - \mathcal{E} y(0)] \\ = \nabla J(v) + 2S^* \mathcal{E}^* \eta(v) = \nabla J(v) + Z(v) \quad . \end{aligned}$$

Remarquons que le calcul de $\nabla J(v) + Z(v)$ nécessite d'appliquer l'opérateur S^* , donc de résoudre un système aux dérivées partielles adjoint.

$$\text{En effet } \nabla J(v) + Z(v) = 2S^* \mathcal{E}^* [\mathcal{E} y(v) - z_d + \eta(v)]$$

et on a vu au § II.5.1 que si $p_1(v)$ est solution de

$$A^* p_1(v) = \mathcal{E}^* \Lambda_{\mathcal{H}} (\mathcal{E} y(v) - z_d + \eta(v))$$

$$\text{alors } \nabla J(v) + Z(v) = 2(B^* p_1(v) + \Lambda_{\mathcal{U}} Nv) \quad .$$

Pour mettre en oeuvre l'algorithme (5.3.2), connaissant U_n et P_n , on calcule $\nabla J(U_n) + Z_n$ comme il vient d'être dit. on en déduit U_{n+1} puis P_{n+1} par (5.3.2).

V.3.3.- Introduction de nouveaux espaces

L'étude de la convergence de la variable primale pourrait être conduite à l'aide des seules notions définies jusqu'ici. Cependant, l'introduction de

quelques notations supplémentaires permettra l'étude de la convergence de la variable duale.

Notons $\tilde{\mathcal{U}} = L^2(\omega, \mathcal{U})$ espace des variables aléatoires à valeurs dans $\mathcal{U}$ dont le carré de la norme est intégrable.

De même $\tilde{\Psi} = L^2(\omega, \Psi)$

et $\tilde{\Lambda} = \{q \in \tilde{\Psi} \mid \tilde{q}(\omega) \in \Lambda \text{ presque sûrement}\}$

$$\tilde{\Phi}(\tilde{v}, \tilde{q}) = E[(\tilde{q}, B(\tilde{v}))_{\Psi}] = (\tilde{q}, B(\tilde{v}))_{\tilde{\Psi}}$$

$$\tilde{\mathcal{U}}'_{ad} = \{\tilde{v} \in \tilde{\mathcal{U}} \mid \tilde{v}(\omega) \in \mathcal{U}'_{ad} \text{ p.s.}\}$$

$$\tilde{\mathcal{U}}_{ad} = \{\tilde{v} \in \tilde{\mathcal{U}} \mid \tilde{v}(\omega) \in \mathcal{U}_{ad} \text{ p.s.}\}.$$

En identifiant $\mathcal{U}_{ad}$ à un sous-ensemble de $\mathcal{U}$ formé de variables aléatoires presque sûrement constantes, $\mathcal{U}_{ad} \subset \tilde{\mathcal{U}}_{ad}$. De même $\Lambda \subset \tilde{\Lambda}$.

$$\text{Posons } L(v, q) = I(v) + (q, B(v))_{\Psi}$$

$$E[\tilde{\Phi}(\tilde{v}, \tilde{q})] = E[J(\tilde{v})] + E[(\tilde{q}, B(\tilde{v}))_{\Psi}].$$

Nous allons comparer les points-cols de $\tilde{\Phi}(v, q)$ et ceux de $L(\tilde{v}, \tilde{q})$ sur $\tilde{\mathcal{U}}_{ad} \times \tilde{\Lambda}$.

Proposition 5.3.3.1.-

Sous l'hypothèse $\underline{J}_3$, $I(\tilde{v})$ est dérivable au sens de Frechet dans $\tilde{\mathcal{U}}$ et $VI(\tilde{v})(\omega) = \nabla J(\tilde{v}(\omega))$.

Démonstration :

$$\begin{array}{ccc} J : v & \longmapsto & J(v) \\ \mathcal{U} & \longrightarrow & \mathbb{R} \end{array} \quad \begin{array}{ccc} I : \tilde{\mathcal{U}} & \longrightarrow & E[J(\tilde{v})] \\ \mathcal{U} & \longrightarrow & \mathbb{R} \end{array}$$

$$J(\tilde{v} + \delta\tilde{v}) - J(\tilde{v}) = \delta J = (\nabla J(\tilde{v}), \delta\tilde{v})_{\mathcal{U}} + \int_{\theta}^1 d\theta (\nabla J(\tilde{v} + \theta\delta\tilde{v}) - \nabla J(\tilde{v}), \delta\tilde{v})_{\mathcal{U}}$$

$$\delta I = \int_{\omega} \delta J \, dp$$

$$\delta I = (\nabla J(\tilde{v}), \delta\tilde{v})_{\mathcal{U}} + \int_{\omega} dp(\omega) \int_{\theta}^1 d\theta (\nabla J(\tilde{v} + \theta\delta\tilde{v}) - \nabla J(\tilde{v}), \delta\tilde{v})_{\mathcal{U}}$$

$$\left| \int_{\omega} dp \int_{\theta}^1 d\theta (\quad) \right| \leq \int_{\omega} dp \frac{M}{2} \|\delta\tilde{v}(\omega)\|_{\mathcal{U}}^2 = \frac{M}{2} \|\delta\tilde{v}\|_{\mathcal{U}}^2 \text{ d'après } \underline{J}_3.$$

On peut donc écrire $\delta I = (\nabla J(\tilde{v}), \delta\tilde{v})_{\mathcal{U}} + R(\tilde{v}, \delta\tilde{v}) \times \|\delta\tilde{v}\|_{\mathcal{U}}$
 où $R(\tilde{v}, \delta\tilde{v}) \xrightarrow{\|\delta\tilde{v}\|_{\mathcal{U}} \rightarrow 0} 0$.

Donc $\nabla J(\tilde{v}(\omega)) = \nabla I(\tilde{v}) (\omega)$.

Considérons à présent le problème de point col

$$L(\tilde{u}, \tilde{q}) \leq L(\tilde{u}, \tilde{p}) \leq L(\tilde{v}, \tilde{p}) \quad \forall (\tilde{v}, \tilde{q}) \in \tilde{\mathcal{U}}_{ad} \times \tilde{\Lambda} .$$

Proposition 5.3.3.2.-

Sous les hypothèses $\underline{J}_2$ et $\underline{J}_3$, si $(\tilde{u}_1, \tilde{p}_1)$ et $(\tilde{u}_2, \tilde{p}_2)$ sont points cols de $L(\tilde{v}, \tilde{q})$, alors $\tilde{u}_1 = \tilde{u}_2$.

Démonstration :

Il suffit de déduire des hypothèses $\underline{J}_2$ et $\underline{J}_3$ (faites sur J) des hypothèses analogues relatives à I .

$$\underline{J}_3 \implies \|\nabla J(\tilde{u}(\omega)) - \nabla J(\tilde{v}(\omega))\|_{\mathcal{U}} \leq M \|\tilde{u}(\omega) - \tilde{v}(\omega)\|_{\mathcal{U}} .$$

En élevant au carré et en intégrant :

$$\|\nabla I(\tilde{u}) - \nabla I(\tilde{v})\|_{\tilde{\mathcal{U}}} \leq M \|\tilde{u} - \tilde{v}\|_{\tilde{\mathcal{U}}} .$$

On opère de même pour $\underline{J}_2$.

Proposition 5.3.3.3.-

Soit (u, p) un point col de $\mathcal{L}(v, q)$. (u, p) est point col de $L(\tilde{v}, \tilde{q})$.

Démonstration :

On sait que $\mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(v, p) \quad \forall v \in \mathcal{U}_{ad} , \forall q \in \Lambda$.

En choisissant $v = \tilde{v}(\omega)$, $q = \tilde{q}(\omega)$, $\tilde{v} \in \tilde{\mathcal{U}}_{ad}$, $\tilde{q} \in \tilde{\Lambda}$ et en prenant l'espérance mathématique dans la double inégalité,

$$L(u, \tilde{q}) \leq L(u, p) \leq L(\tilde{v}, p) \quad \forall (\tilde{v}, \tilde{q}) \in \tilde{\mathcal{U}}_{ad} \times \tilde{\Lambda} .$$

V. 4.- Convergence de l'algorithme d'Uzawa stochastique

On complète la définition (5.3.1) par le choix de ρ_n

$$\sum_n \rho_n = +\infty , \quad \sum_n \rho_n^2 < +\infty .$$

Nous allons établir un théorème de convergence de la variable primale, puis un théorème de convergence de la variable duale.

Donc $V(\hat{V}(u)) = V(\hat{V}(u))$.

Considérons à présent le problème de point col

$$L(u, p) \leq L(\hat{u}, p) \leq L(\hat{u}, p) \in \mathcal{V}(\hat{u}, p) \in \mathcal{V}(\hat{u}, p) \times \mathcal{V}$$

et u_3 si (u_1, p_1) et (u_2, p_2) sont points

démonstration :

Soit $(u, p) \in \mathcal{V}(\hat{u}, p)$ et $(u, p) \in \mathcal{V}(\hat{u}, p)$.

$$\|u - \hat{u}\| \leq \|u - \hat{u}\| \leq \|u - \hat{u}\|$$

on a donc $\|u - \hat{u}\| \leq \|u - \hat{u}\|$.

$$\|u - \hat{u}\| \leq \|u - \hat{u}\| \leq \|u - \hat{u}\|$$

on a donc $\|u - \hat{u}\| \leq \|u - \hat{u}\|$.

est (u, p) un point col de $\mathcal{V}(\hat{u}, p)$ est point col de $\mathcal{V}(\hat{u}, p)$.

démonstration :

Soit $(u, p) \in \mathcal{V}(\hat{u}, p)$ et $(u, p) \in \mathcal{V}(\hat{u}, p)$.

$$\|u - \hat{u}\| \leq \|u - \hat{u}\| \leq \|u - \hat{u}\|$$

mathématique dans la double

$$L(u, p) \leq L(\hat{u}, p) \leq L(\hat{u}, p) \in \mathcal{V}(\hat{u}, p) \in \mathcal{V}(\hat{u}, p) \times \mathcal{V}$$

Soit $(u, p) \in \mathcal{V}(\hat{u}, p)$ et $(u, p) \in \mathcal{V}(\hat{u}, p)$.

démonstration (5.3.1) par le choix de

$$\|u - \hat{u}\| \leq \|u - \hat{u}\| \leq \|u - \hat{u}\|$$

5.4.1.- Convergence de la variable primale

Théorème 5.4.1.-

Soit u la solution du problème primal. Sous les hypothèses $\underline{J}_1, \underline{J}_2, \underline{J}_3, \underline{Q}, \underline{S}_1, \underline{S}_2$ il existe une sous-suite $\{U_{n_\ell}\}$ telle que $U_{n_\ell} - Z_{n_\ell} \xrightarrow{\ell \rightarrow \infty} u$ en moyenne quadratique. Par suite $\limsup_{\ell} \|U_{n_\ell} - u\|_{\mathcal{U}}^2 \leq B^2$

Démonstration :

Soit u la solution (unique) du problème primal et p une solution du problème dual. $p = p_\Lambda (p + \rho_n B(u))$.

Posons $R_n = P_n - p$

$$(5.4.1) \quad \|R_{n+1}\|_{\Psi}^2 \leq \|R_n + \rho_n (B(U_n) - B(u))\|_{\Psi}^2 \quad (\text{proposition 1.5.1.1})$$

$$U_n - Z_n \text{ minimise } J(v) + (P_n, B(v))_{\Psi} \text{ sur } \mathcal{U}_{ad} \iff$$

$$(5.4.2) \quad (\nabla J(U_n - Z_n), v - (U_n - Z_n))_{\mathcal{U}} + (P_n, B(v) - B(U_n - Z_n))_{\Psi} \geq 0 \quad \forall v \in \mathcal{U}_{ad}$$

$$\implies \text{ avec } v = u$$

$$(5.4.3) \quad (\nabla J(U_n - Z_n), u - (U_n - Z_n))_{\mathcal{U}} + (P_n, B(u) - B(U_n - Z_n))_{\Psi} \geq 0.$$

La deuxième inégalité de (5.1.3) s'écrit :

$$(5.4.4) \quad (\nabla J(u), v - u)_{\mathcal{U}} + (p, B(v) - B(u))_{\Psi} \geq 0 \quad \forall v \in \mathcal{U}_{ad}$$

$$\implies \text{ avec } v = U_n - Z_n$$

$$(5.4.5) \quad (\nabla J(u), U_n - Z_n - u)_{\mathcal{U}} + (p, B(U_n - Z_n) - B(u))_{\Psi} \geq 0$$

$$\implies (5.4.5) - (5.4.3) :$$

$$(5.4.6) \quad (\nabla J(U_n - Z_n) - \nabla J(u), U_n - Z_n - u)_{\mathcal{U}} + (P_n - p, B(U_n - Z_n) - B(u))_{\Psi} \leq 0$$

$$(5.4.6) \text{ et } \underline{J}_2 \implies$$

$$(5.4.7) \quad - (P_n - p, B(U_n - Z_n) - B(u))_{\Psi} \geq m \|U_n - Z_n - u\|_{\mathcal{U}}^2$$

$$(5.4.1) \implies$$

$$(5.4.8) \quad \|R_{n+1}\|_{\Psi}^2 \leq \|R_n\|_{\Psi}^2 + 2\rho_n (R_n, B(U_n) - B(u))_{\Psi} + \rho_n^2 \|B(U_n) - B(u)\|_{\Psi}^2.$$

D'après les hypothèses sur le bruit Z_n ,

$$E [B(U_n) | P_n] = E [B(U_n - Z_n + Z_n) | P_n] = B(U_n - Z_n)$$

$$\begin{aligned} E [(R_n, B(U_n) - B(u))_\Psi] &= (R_n, B(U_n) - B(u))_\Psi^\sim \\ &= (R_n, B(U_n) - B(U_n - Z_n))_\Psi^\sim + (R_n, B(U_n - Z_n))_\Psi^\sim \\ (R_n, B(U_n) - B(U_n - Z_n))_\Psi^\sim &= E \{ E [(R_n, B(U_n) - B(U_n - Z_n))_\Psi | P_n] \} \\ &= E \{ (R_n, E [B(U_n) | P_n] - B(U_n - Z_n))_\Psi \} \\ &= 0 \end{aligned}$$

$$\text{Donc } (R_n, B(U_n) - B(u))_\Psi^\sim = (R_n, B(U_n - Z_n) - B(u))_\Psi^\sim.$$

On montre de même que

$$\|B(U_n) - B(u)\|_\Psi^2 = \|B(U_n) - B(U_n - Z_n)\|_\Psi^2 + \|B(U_n - Z_n) - B(u)\|_\Psi^2.$$

Donc (5.4.8) devient

$$\begin{aligned} \|R_{n+1}\|_\Psi^2 &\leq \|R_n\|_\Psi^2 + 2\rho_n (R_n, B(U_n - Z_n) - B(u))_\Psi^\sim + \rho_n^2 [\|B(U_n) - B(U_n - Z_n)\|_\Psi^2 \\ &\quad + \|B(U_n - Z_n) - B(u)\|_\Psi^2] \end{aligned}$$

(5.4.7) $\implies$

$$\|R_{n+1}\|_\Psi^2 \leq \|R_n\|_\Psi^2 - 2\rho_n - m \|U_n - Z_n - u\|_{\mathcal{U}}^2 + \rho_n^2 C^2 [\|Z_n\|_{\mathcal{W}}^2 + \|U_n - Z_n - u\|_{\mathcal{U}}^2]$$

ou :

$$(5.4.9) \quad \|R_{n+1}\|_\Psi^2 \leq \|R_n\|_\Psi^2 + (\rho_n^2 C^2 - 2m\rho_n) \|U_n - Z_n - u\|_{\mathcal{U}}^2 + C^2 \rho_n^2 \|Z_n\|_{\mathcal{W}}^2$$

$$\rho_n \rightarrow 0 \implies 2m\rho_n - \rho_n^2 C^2 > 0 \text{ pour } n \text{ assez grand}$$

$$\|Z(q)\|_{\mathcal{W}}^2 \leq B^2 \implies \|Z_n\|_{\mathcal{W}}^2 \leq B^2.$$

$$\text{D'où } \|R_n\|_\Psi^2 - \|R_{n+1}\|_\Psi^2 \geq (2m\rho_n - \rho_n^2 C^2) \|U_n - Z_n - u\|_{\mathcal{U}}^2 - \rho_n^2 C^2 B^2.$$

Comme on a supposé que $\sum_n \rho_n^2 < +\infty$, $\|R_n\|_{\mathcal{W}}^2$ a une limite d'après la proposition (1.5.2.2) et par suite (prop. (1.5.2.3))

$$\sum_n (2m\rho_n - \rho_n^2 C^2) \|U_n - Z_n - u\|_{\mathcal{U}}^2 < +\infty$$

et il existe une sous-suite telle que $\|U_{n_\ell} - Z_{n_\ell} - u\|_{\mathcal{U}}^\sim \xrightarrow{\ell \rightarrow \infty} 0$

$$\|U_{n_\ell} - u\|_{\mathcal{U}}^2 = \|U_{n_\ell} - Z_{n_\ell} - u\|_{\mathcal{U}}^2 + \|Z_{n_\ell}\|_{\mathcal{U}}^2 \text{ car } U_{n_\ell} - Z_{n_\ell} \text{ est mesurable}$$

$$\text{par rapport à } F_{n_\ell} \implies \limsup_{\ell} \|U_{n_\ell} - u\|_{\mathcal{U}}^2 \leq B^2.$$

V.4.2.- Convergence de la variable duale

Théorème 5.4.2.-

L'ensemble $\{P'\}$ des points d'adhérence faible de la suite $\{P_{n_\ell}\}$ ($\{n_\ell\}$ définie au théorème 5.4.1) sont tels que (u, P') soit point-col de $L(\hat{v}, \hat{q})$.

Démonstration :

Prouvons d'abord que $(B(U_n - Z_n), P_{n+1} - P_n)_\Psi$ est le terme général ≥ 0 d'une série convergente.

Si u_q est défini par $(\nabla J(u_q), v - u_q)_\mathcal{U} + (q, B(v) - B(u_q))_\Psi \geq 0 \quad \forall v \in \mathcal{U}_{ad}$, en posant $g(q) = \mathcal{L}(u_q, q)$, la 2ème inégalité de (5.1.4) prouve que

$$\nabla g(q) = B(u_q)$$

$$g(P_{n+1}) - g(P_n) = (B(U_{P_n}), P_{n+1} - P_n)_\Psi + \int_0^1 \frac{d\theta}{\theta} (B(U_{P_n + \theta(P_{n+1} - P_n)}) - B(U_{P_n}), P_n + \theta(P_{n+1} - P_n) - P_n)_\Psi.$$

En utilisant la définition de u_q d'une part avec $q = P_n + \theta(P_{n+1} - P_n)$ $v = U_{P_n}$ et d'autre part avec $q = P_n$, $v = U_{P_n + \theta(P_{n+1} - P_n)}$

$$\begin{aligned} -C \|U_{P_n + \theta(P_{n+1} - P_n)} - U_{P_n}\|_{\mathcal{U}} \|P_n + \theta(P_{n+1} - P_n) - P_n\|_\Psi &\leq \\ (B(U_{P_n + \theta(P_{n+1} - P_n)}) - B(U_{P_n}), P_n + \theta(P_{n+1} - P_n) - P_n)_\Psi &\leq \\ \leq -(\nabla J(U_{P_n + \theta(P_{n+1} - P_n)}) - \nabla J(U_{P_n}), U_{P_n + \theta(P_{n+1} - P_n)} - U_{P_n})_{\mathcal{U}} &\leq \\ \leq -m \|U_{P_n + \theta(P_{n+1} - P_n)} - U_{P_n}\|_{\mathcal{U}}^2 \leq 0 \end{aligned}$$

$$\implies \|U_{P_n + \theta(P_{n+1} - P_n)} - U_{P_n}\|_{\mathcal{U}} \leq \frac{C\theta}{m} \|P_{n+1} - P_n\|_\Psi.$$

$$\begin{aligned} \text{Donc } 0 &\geq \int_0^1 d\theta \ C \|U_{P_n + \theta(P_{n+1} - P_n)} - U_{P_n}\|_{\mathcal{U}} \|P_{n+1} - P_n\|_\Psi \\ &\geq -\frac{C^2}{2m} \|P_{n+1} - P_n\|_\Psi^2. \end{aligned}$$

Donc (5.4.10) $g(P_{n+1}) - g(P_n) = (B(U_n - Z_n), P_{n+1} - P_n)_\Psi$

$$= \theta_n^2 \frac{C^2}{2m} \|P_{n+1} - P_n\|_\Psi^2$$

θ_n^2 aléatoire tel que $0 \leq \theta_n^2 \leq 1$.

Nous allons maintenant prendre l'espérance mathématique après avoir relié

$(B(U_n), P_{n+1} - P_n)_\Psi$ et $(B(U_n - Z_n), P_{n+1} - P_n)_\Psi$

$P_\Lambda(P_n + \rho_n B(U_n - Z_n))$ est mesurable par rapport à $P_n \Rightarrow$

$$0 \leq (P_{n+1} - P_n, B(U_n) - B(U_n - Z_n))_\Psi = (P_{n+1} - P_\Lambda(P_n + \rho_n B(U_n - Z_n)), B(U_n) - B(U_n - Z_n))_\Psi$$

$$\leq \rho_n \|B(U_n) - B(U_n - Z_n)\|_\Psi^2 \leq \rho_n C^2 \|Z_n\|^2 \leq \rho_n C^2 B^2$$

$$\|P_{n+1} - P_n\|_\Psi^2 \leq (P_{n+1} - P_n, \rho_n B(U_n))_\Psi = (P_{n+1} - P_n, \rho_n B(U_n - Z_n))_\Psi + (P_{n+1} - P_n, \rho_n (B(U_n) - B(U_n - Z_n)))_\Psi$$

$$(5.4.12) \quad \|P_{n+1} - P_n\|_\Psi^2 \leq \rho_n (P_{n+1} - P_n, B(U_n - Z_n))_\Psi + \rho_n^2 B^2 C^2$$

ou $\|P_{n+1} - P_n\|_\Psi^2 = \theta_n \rho_n (P_{n+1} - P_n, B(U_n - Z_n))_\Psi + \theta_n \rho_n^2 B^2 C^2$

$$0 \leq \theta_n \leq 1$$

En prenant l'espérance mathématique dans (5.4.10)

$$(5.4.13) \quad E[g(P_{n+1})] - E[g(P_n)] = \rho_n (P_{n+1} - P_n, B(U_n - Z_n))_\Psi \left[\frac{1}{\rho_n} - \rho_n \frac{C^2}{2m} \right] - \theta'_n \rho_n^2 B^2 C^2$$

$$0 \leq \theta'_n \leq 1$$

$$\begin{aligned} (P_{n+1} - P_n, B(U_n - Z_n))_\Psi &= E\{E[(P_{n+1} - P_n, B(U_n - Z_n))_\Psi | P_n]\} = \\ &= E\{(E(P_{n+1} | P_n) - P_n, B(U_n - Z_n))_\Psi\} \\ &= (P_\Lambda(P_n + \rho_n B(U_n - Z_n)) - P_n, B(U_n - Z_n))_\Psi \geq 0 \end{aligned}$$

Donc (5.4.13) s'écrit $E[g(P_{n+1})] - E[g(P_n)] = v_n + w_n$

où $\sum_n |w_n| < +\infty$ $v_n \geq 0$ car $\left[\frac{1}{\rho_n} - \theta'_n \frac{C^2}{2m} \right] > 0$ pour n assez grand.

D'après la proposition 1.5.2.2 $\sum v_n < +\infty$

et donc aussi (5.4.14) $\sum_n (P_{n+1} - P_n, B(U_n - Z_n))_\Psi < +\infty$.

Nous allons à présent montrer que $(R_{n_\ell}, B(U_{n_\ell} - Z_{n_\ell}))_{\tilde{\Psi}} \xrightarrow{\ell \rightarrow \infty} 0$.

Considérons la suite $\{n_\ell\}$ extraite au théorème (5.4.1) $\|R_n\|_{\tilde{\Psi}}^2$ a une limite. Donc $\{R_n\}$ est bornée dans $\tilde{\Psi}$.

On peut extraire de $\{n_\ell\}$ une suite, soit $\{n'\}$ telle que $R_{n'} \rightarrow R$ dans $\tilde{\Psi}$ faible.

D'après le théorème (5.4.1) $U_{n'} - Z_{n'} \rightarrow u$ dans $\mathcal{U}$ fort
et $\|B(U_{n'} - Z_{n'}) - B(u)\|_{\tilde{\Psi}}^2 \rightarrow 0$

Il résulte alors de (5.4.7) que

$$(5.4.15) \quad (R_{n'}, B(U_{n'} - Z_{n'}))_{\tilde{\Psi}} \rightarrow (R, B(u))_{\tilde{\Psi}}.$$

$P_{n'} \in \tilde{\Lambda}$ convexe fermé, donc faiblement fermé.

Donc $R + p \in \tilde{\Lambda} \Rightarrow (5.4.16) \quad (R, B(u))_{\tilde{\Psi}} \leq 0$ (analogue à 5.1.4)

D'après (5.4.14), $(R_{n+1} - R_n, B(U_n - Z_n))_{\tilde{\Psi}} \xrightarrow{n \rightarrow \infty} 0$

et avec (5.4.15),

$$(5.4.17) \quad (R_{n'+1}, B(U_{n'} - Z_{n'}))_{\tilde{\Psi}} \rightarrow (R, B(u))_{\tilde{\Psi}}.$$

Comme conséquence des propriétés de l'opérateur de projection P_Λ :

$$(P_{n'} + \rho_{n'} B(U_{n'})) - P_{n'+1}, z - P_{n'+1})_{\tilde{\Psi}} \leq 0 \quad \forall z \in \Lambda$$

avec $z = P$:

$$(5.4.18) \quad - (R_{n'} - R_{n'+1} + \rho_{n'} B(U_{n'}), R_{n'+1})_{\tilde{\Psi}} \leq 0.$$

Soit $\rho_{n'} (B(U_{n'}), R_{n'+1})_{\tilde{\Psi}} \geq (R_{n'+1}, R_{n'+1} - R_{n'})_{\tilde{\Psi}} \geq$

$$\geq \frac{1}{2} \|R_{n'+1}\|_{\tilde{\Psi}}^2 - \frac{1}{2} \|R_{n'}\|_{\tilde{\Psi}}^2.$$

D'où

$$(5.4.19) \quad (B(U_{n'}), R_{n'+1})_{\tilde{\Psi}} \geq \frac{1}{2\rho_{n'}} [\|R_{n'+1}\|_{\tilde{\Psi}}^2 - \|R_{n'}\|_{\tilde{\Psi}}^2]$$

$$(5.4.12) \Rightarrow \frac{1}{\rho_n} \|R_{n+1} - R_n\|_{\tilde{\Psi}}^2 \leq (B(U_n - Z_n), P_{n+1} - P_n)_{\tilde{\Psi}} + \rho_n B^2 C^2 \xrightarrow{n \rightarrow \infty} 0$$

$$\|R_{n+1}\|_{\tilde{\Psi}}^2 \leq 2 [\|R_{n+1} - R_n\|_{\tilde{\Psi}}^2 + \|R_n\|_{\tilde{\Psi}}^2]$$

$$\Rightarrow \frac{1}{\rho_n} \|R_{n+1} - R_n\|_{\tilde{\Psi}}^2 \geq \frac{1}{2\rho_n} [\|R_{n+1}\|_{\tilde{\Psi}}^2 - \|R_n\|_{\tilde{\Psi}}^2]$$

c'est-à-dire (5.4.20)
$$\frac{1}{2\rho_n} \left[\|R_{n+1}\|_{\tilde{\Psi}}^2 - \|R_n\|_{\tilde{\Psi}}^2 \right] \xrightarrow{n \rightarrow \infty} 0$$

(5.4.21)
$$(B(U_n - Z_n), R_{n+1})_{\tilde{\Psi}} = (B(U_n), R_{n+1})_{\tilde{\Psi}} - (B(U_n) - B(U_n - Z_n), R_{n+1})_{\tilde{\Psi}}$$

(5.4.11) s'écrit

$$0 \leq (R_{n+1} - R_n, B(U_n) - B(U_n - Z_n))_{\tilde{\Psi}} \leq \rho_n B^2 C^2.$$

Mais $(R_n, B(U_n) - B(U_n - Z_n))_{\tilde{\Psi}} = 0$.

Donc :

(5.4.22)
$$0 \leq (R_{n+1}, B(U_n) - B(U_n - Z_n))_{\tilde{\Psi}} \leq \rho_n B^2 C^2.$$

En portant (5.4.19) et (5.4.22) dans (5.4.21)

$$(B(U_{n'} - Z_{n'}), R_{n'+1})_{\tilde{\Psi}} \geq \frac{1}{2\rho_{n'}} \left[\|R_{n'+1}\|_{\tilde{\Psi}}^2 - \|R_{n'}\|_{\tilde{\Psi}}^2 \right] - \rho_{n'} B^2 C^2.$$

Il résulte alors de (5.4.17) et (5.4.20) que

(5.4.23)
$$(R, B(u))_{\tilde{\Psi}} \geq 0$$

ce qui comparé à (5.4.15) donne

(5.4.24)
$$(R, B(u))_{\tilde{\Psi}} = 0.$$

Donc 0 est le seul point d'accumulation de la suite $(R_n, B(U_n - Z_n))_{\tilde{\Psi}}$.
D'où

(5.4.25)
$$(R_{n_\ell}, B(U_{n_\ell} - Z_{n_\ell}))_{\tilde{\Psi}} \xrightarrow{\ell \rightarrow \infty} 0$$

Il en résulte que l'on a

(5.4.26)
$$(P_{n_\ell}, B(U_{n_\ell} - Z_{n_\ell}))_{\tilde{\Psi}} \xrightarrow{\ell \rightarrow \infty} (P, B(u))_{\tilde{\Psi}}.$$

Par définition de p , $\mathcal{E}(u, q) \leq \mathcal{E}(u, p) \quad \forall q \in \Lambda$

$$\iff (q, B(u))_{\tilde{\Psi}} \leq (p, B(u))_{\tilde{\Psi}} \quad \forall q \in \Lambda$$

Donc $(\tilde{q}, B(u))_{\tilde{\Psi}} \leq (p, B(u))_{\tilde{\Psi}} \quad \forall \tilde{q} \in \tilde{\Lambda}$

et (5.4.26) s'écrit

(5.4.27)
$$(P_{n_\ell}, B(U_{n_\ell} - Z_{n_\ell}))_{\tilde{\Psi}} \xrightarrow{\ell \rightarrow \infty} \sup_{\tilde{q} \in \tilde{\Lambda}} (\tilde{q}, B(u))_{\tilde{\Psi}}.$$

Soit une sous-suite $\{P_{n_\ell}\}$ extraite de $\{P_n\}$ telle que

$$P_{n_\ell} \longrightarrow P' \in \tilde{\Lambda} \quad \text{dans } \tilde{\Psi} \text{ faible}$$

$$U_{n_\ell} - Z_{n_\ell} \longrightarrow u \text{ dans } \tilde{\mathcal{U}} \text{ fort.}$$

D'où

$$(P_{n_\ell}, B(U_{n_\ell} - Z_{n_\ell}))_{\tilde{\Psi}} \longrightarrow (P', B(u))_{\tilde{\Psi}} = \sup_{\tilde{q} \in \tilde{\Lambda}} (\tilde{q}, B(u))_{\tilde{\Psi}}$$

c'est-à-dire

$$(5.4.28) \quad (\tilde{q} - P', B(u))_{\tilde{\Psi}} \leq 0 \quad \forall \tilde{q} \in \tilde{\Lambda}.$$

Soit

$$(5.4.29) \quad L(u, \tilde{q}) \leq L(u, P') \quad \forall \tilde{q} \in \tilde{\Lambda}.$$

Les hypothèses faites sur J impliquent :

$$\nabla J(U_{n_\ell} - Z_{n_\ell}) \longrightarrow \nabla J(u) \text{ dans } \tilde{\mathcal{U}} \text{ faible.}$$

$$(5.4.2) \implies$$

$$(\nabla J(U_{n_\ell} - Z_{n_\ell}), \tilde{v} - (U_{n_\ell} - Z_{n_\ell}))_{\tilde{\mathcal{U}}} + (P_{n_\ell}, B(\tilde{v}) - B(U_{n_\ell} - Z_{n_\ell}))_{\tilde{\Psi}} \geq 0 \quad \forall \tilde{v} \in \tilde{\mathcal{U}}_{\text{ad}}.$$

En passant à la limite :

$$(\nabla J(u), \tilde{v} - u)_{\tilde{\mathcal{U}}} + (P', B(\tilde{v}) - B(u))_{\tilde{\Psi}} \geq 0 \quad \forall \tilde{v} \in \tilde{\mathcal{U}}_{\text{ad}}$$

$$\text{c'est-à-dire :} \quad L(u, P') \leq L(\tilde{v}, P') \quad \forall \tilde{v} \in \tilde{\mathcal{U}}_{\text{ad}}.$$

V.4.3.- Comparaison avec les résultats antérieurs

L'algorithme d'Uzawa envisagé sur un plan stochastique n'a été l'objet à notre connaissance, d'aucune publication jusqu'à ce jour.

Il n'en est pas de même pour les algorithmes déterministes qui en dérivent.

Avec un choix de ρ_n différent ($\rho_n \geq \delta > 0$).

A. Bensoussan, J.L. Lions, R. Temam [1] démontrent des théorèmes déterministes comparables aux Théorèmes (5.4.1) et (5.4.2).

Avec l'hypothèse $\rho_n \longrightarrow 0 \quad \sum_n \rho_n = +\infty$ (algorithme à petits pas),

A. Auslender [2] démontre la convergence de la variable primale dans le cas déterministe.

En revanche, nous n'avons pas connaissance de résultat déterministe analogue au théorème (5.4.2) pour l'algorithme à petit pas.

Un tel résultat est fourni par le théorème (5.4.2) évidemment encore valable dans le cas déterministe.

V.5.- Convergence de l'algorithme d'Arrow-Hurwicz stochastique

On considère l'algorithme (5.3.2) où on va choisir $\rho_n^{(1)} = \rho_n$
 $\rho_n^{(2)} = k \rho_n$ vérifiant :

$$\rho_n > 0 \quad \sum_n \rho_n = +\infty, \quad \sum_n \rho_n^2 < +\infty, \quad \rho_n \leq \frac{m}{M^2}.$$

V.5.1.- Convergence de la variable primale

Théorème 5.5.1.-

Soit u la solution du problème primal. Sous les hypothèses $\underline{J}_1$, $\underline{J}_2$, $\underline{J}_3$, $\underline{Q}$, $\underline{S}_1$, $\underline{S}_2$, il existe une sous-suite $\{U_{n_\ell}\}$ telle que $U_{n_\ell} \xrightarrow{\ell \rightarrow \infty} u$ en moyenne quadratique.

Démonstration :

D'après $\underline{J}_2$ et $\underline{J}_3$, on a

$$(5.5.1) \quad \|v - u - \rho (\nabla J(v) - \nabla J(u))\|_{\mathcal{U}}^2 \leq [1 - 2\rho m + \rho^2 M^2] \|v - u\|_{\mathcal{U}}^2.$$

Comme $1 - \rho m \geq 1 - 2\rho m + \rho^2 M^2$ dès que $\rho \leq \frac{m}{M^2}$

$$(5.5.2) \quad \|v - u - \rho (\nabla J(v) - \nabla J(u))\|_{\mathcal{U}}^2 \leq (1 - \rho m) \|v - u\|_{\mathcal{U}}^2.$$

Soit (u, p) point-col de $\mathcal{L}(v, q)$

$$(5.5.3) \quad \begin{cases} u = u - \rho_n (\nabla J(u) + B^* p) \\ p = P_\Lambda(p - k \rho_n B u) \end{cases}.$$

Posons $W_n = U_n - u$, $R_n = P_n - p$

$$W_{n+1} = W_n - \rho_n (\nabla J(U_n) - \nabla J(u) + B^* R_n + Z_n)$$

$$\|R_{n+1}\|_\Psi^2 = \|P_\Lambda(P_n + k \rho_n B U_{n+1}) - P_\Lambda(p - k \rho_n B u)\|_\Psi^2$$

$$(5.5.4) \quad \|R_{n+1}\|_{\Psi}^2 \leq \|R_n + k\rho_n BW_{n+1}\|_{\Psi}^2 = \|R_n\|_{\Psi}^2 + 2k\rho_n (R_n, BW_{n+1})_{\Psi} + \rho_n^2 k^2 \|BW_{n+1}\|_{\Psi}^2$$

$$(5.5.5) \quad \|W_{n+1}\|_{\mathcal{U}}^2 = (W_n - \rho_n (\nabla J(U_n) - \nabla J(u)), W_{n+1})_{\mathcal{U}} - \rho_n (B^* R_n, W_{n+1})_{\mathcal{U}} - \rho_n (Z_n, W_{n+1})_{\mathcal{U}}$$

$$\begin{aligned} \mathbb{E} [-\rho_n (Z_n, W_{n+1})_{\mathcal{U}} | U_n] &= \mathbb{E} [-\rho_n (Z_n, W_n - \rho_n (\nabla J(U_n) - \nabla J(u)) + B^* R_n + Z_n)_{\mathcal{U}} | U_n] \\ &= \underbrace{\rho_n^2 \mathbb{E} [\|Z_n\|_{\mathcal{U}}^2 | U_n]}_{\leq B^2} + \underbrace{\mathbb{E} [-\rho_n (Z_n, W_n - \rho_n (\nabla J(U_n) - \nabla J(u)) + B^* R_n)_{\mathcal{U}} | U_n]}_{= 0 \text{ (mesurabilité du 2ème terme par rapport à } U_n)} \end{aligned}$$

Donc :

$$(5.5.6) \quad -\rho_n (Z_n, W_{n+1})_{\mathcal{U}} \leq \rho_n^2 B^2$$

$$(5.5.5) \text{ et } (5.5.6) \Rightarrow \|W_{n+1}\|_{\mathcal{U}}^2 \leq (W_n - \rho_n (\nabla J(U_n) - \nabla J(u)), W_{n+1})_{\mathcal{U}} - \rho_n (B^* R_n, W_{n+1})_{\mathcal{U}} + \rho_n^2 B^2.$$

Puisque $\rho_n \leq \frac{m}{M^2}$, (5.5.2) peut être utilisé, soit en posant

$$\beta_n^2 = 1 - \rho_n m, \quad \beta_n > 0$$

$$(5.5.7) \quad \|W_{n+1}\|_{\mathcal{U}}^2 \leq \beta_n \|W_n\|_{\mathcal{U}}^2 \|W_{n+1}\|_{\mathcal{U}}^2 - \rho_n (B^* R_n, W_{n+1})_{\mathcal{U}} + \rho_n^2 B^2$$

$$(5.5.4) \Rightarrow$$

$$(5.5.8) \quad \|R_{n+1}\|_{\Psi}^2 \leq \|R_n\|_{\Psi}^2 + 2k\rho_n (B^* R_n, W_{n+1})_{\mathcal{U}} + k^2 \rho_n^2 \|B^* W_{n+1}\|_{\Psi}^2.$$

$$\text{D'où } \|R_{n+1}\|_{\Psi}^2 \leq \|R_n\|_{\Psi}^2 + 2k \left[\beta_n \|W_n\|_{\mathcal{U}}^2 \|W_{n+1}\|_{\mathcal{U}}^2 - \|W_{n+1}\|_{\mathcal{U}}^2 + \rho_n^2 B^2 \right] + \rho_n^2 k^2 C^2 \|W_{n+1}\|_{\mathcal{U}}^2$$

$$(5.5.9) \quad \|R_{n+1}\|_{\Psi}^2 \leq \|R_n\|_{\Psi}^2 + k \left[\beta_n \|W_n\|_{\mathcal{U}}^2 - \beta_{n+1} \|W_{n+1}\|_{\mathcal{U}}^2 \right] + 2k \rho_n^2 B^2 + \left[\rho_n^2 k^2 C^2 - 2k + k(\beta_n + \beta_{n+1}) \right] \|W_{n+1}\|^2.$$

$$\text{En posant } \tau_n = \|R_n\|_{\Psi}^2 + k \beta_n \|W_n\|_{\mathcal{U}}^2, \quad (5.5.9) \iff$$

$$(5.5.10) \quad \tau_n - \tau_{n+1} \geq \left[k(2 - (\beta_n + \beta_{n+1})) - \rho_n^2 k^2 C^2 \right] \|W_{n+1}\|_{\mathcal{U}}^2 - 2k \rho_n^2 B^2.$$

A fortiori, en posant $\gamma_n = k(1 - \beta_n) - \rho_n^2 k^2 C^2$

$$(5.5.11) \quad \tau_n - \tau_{n+1} \geq \gamma_n \cdot \|W_{n+1}\|_{\mathcal{U}}^2 - 2k\rho_n^2 B^2.$$

Puisque $\rho_n \longrightarrow 0$ $\gamma_n \sim \frac{m}{2} \rho_n$ lorsque $n \longrightarrow +\infty$.

Donc (5.5.12) $\sum_n \gamma_n = +\infty$ et $\gamma_n > 0$ pour n assez grand.

D'après la proposition (1.5.2.2), τ_n a une limite

$$\text{et } \sum_n \gamma_n \|W_{n+1}\|_{\mathcal{U}}^2 < +\infty.$$

D'après la proposition (1.5.2.3), 0 est point d'adhérence de $\{W_n\}$

i.e. Il existe un sous-suite U_{n_ℓ} telle que

$$U_{n_\ell} \xrightarrow[n \rightarrow +\infty]{} u \quad \text{en moyenne quadratique.}$$

V.5.2.- Comparaison avec les résultats antérieurs - Problèmes ouverts concernant la convergence de la variable duale

Comme pour la méthode d'Uzawa, nous n'avons pas connaissance de publications concernant la méthode d'Arrow-Hurwicz stochastique. Il est cependant utile de comparer les résultats obtenus aux résultats connus pour les algorithmes déterministes :

Algorithme à petit pas : $\rho_n \longrightarrow 0$ $\sum_n \rho_n = +\infty$.

La convergence de l'algorithme déterministe est donnée par A. Auslender [2], mais nous ne connaissons pas de résultat déterministe concernant la convergence de la variable duale.

Algorithme à grand pas : $\rho_n \geq \delta > 0$.

La convergence de l'algorithme déterministe est donnée par J.L. Lions, R. Glowinski, R. Trémoières [1] dans le cas quadratique et (5.5.1) en permet la généralisation pour une fonctionnelle strictement convexe. On a alors convergence dans $\mathcal{U}$ de toute la suite U_n .

Nous ne connaissons pas de résultat déterministe concernant la convergence de la variable duale.

Reprenant une idée de A. Bensoussan, J.L. Lions, R. Tremam [1] développée dans le cadre de l'algorithme d'Uzawa, démontrons le

Théorème 5.5.2.-

La suite $\{p_n\}$ est bornée et toute suite extraite converge faiblement vers p' tel que (u, p') soit point-col de $\mathcal{L}(v, q)$.

Démonstration :

$u_n \longrightarrow u$ dans $\mathcal{U}_{\text{fort}}$. Donc $Bu_n \longrightarrow Bu$ dans fort . La suite $\{r_n\}$ est bornée, il existe donc une sous-suite $\{r_{n'}\}$ telle que $r_{n'} \xrightarrow{n' \rightarrow +\infty} r$ dans Ψ faible. $p_{n'} = r_{n'} + p \longrightarrow r+p$ dans Ψ faible.

Λ est convexe fermé, donc faiblement fermé.

Donc $r + p \in \Lambda$.

$$(q-p, B(u))_{\Psi} \leq 0 \quad \forall q \in \Lambda \implies$$

$$(5.5.13) \quad (r, B(u))_{\Psi} \leq 0.$$

Grâce à la convergence forte de $B(u_n)$ et la convergence faible de $r_{n'}$,

$$(5.5.14) \quad (B(u_n), r_{n'})_{\Psi} \xrightarrow{n' \rightarrow +\infty} (B(u), r)_{\Psi}$$

Comme conséquence des propriétés de l'opérateur P_{Λ}

$$(5.5.15) \quad (P_{n'-1} + \rho_{n'-1} B(u_{n'}) - P_{n'}, z - P_{n'})_{\Psi} \leq 0 \quad \forall z \in \Lambda$$

Avec $z = p$:

$$(5.5.16) \quad \rho_{n'-1} (Bu_{n'}, r_{n'})_{\Psi} \geq (r_{n'}, r_{n'} - r_{n'-1})_{\Psi} \geq \frac{1}{2} [\|r_{n'}\|_{\Psi}^2 - \|r_{n'-1}\|_{\Psi}^2]$$

$$\{\|r_n\|_{\Psi}^2\} \text{ converge } \rho_{n'-1} \geq \delta > 0.$$

Donc par passage à la limite

$$(5.5.17) \quad (r, B(u))_{\Psi} \geq 0.$$

Par comparaison avec (5.5.13)

$$(5.5.18) \quad (r, B(u))_{\Psi} = 0.$$

On peut maintenant achever la démonstration comme A; Bensoussan, J.L. Lions, R. Tenam [1] p. 151-152.

Le problème pour trouver dans le cas stochastique un résultat analogue au théorème (5.4.2) est donc le passage de $\rho_n \geq \delta > 0$ à $\rho_n \longrightarrow 0$, $\sum \rho_n = +\infty$ et non pas le passage du cas déterministe au cas stochastique. Nous avons résolu cette question pour l'algorithme d'Uzawa, mais le problème reste ouvert pour l'algorithme d'Arrow-Hurwicz.

6

applications

numériques

VI.1.- Une application en identification des systèmes aux dérivées partielles

VI.1.1.- Le modèle

Le système est défini par

$$\left\{ \begin{array}{l} \frac{\partial y}{\partial t} - \frac{\partial}{\partial x} \left[A(x) \frac{\partial y}{\partial x} \right] = f \quad \text{sur } (0,1) \times (0,T) \\ \frac{\partial y}{\partial x}(0,t) = \frac{\partial y}{\partial x}(1,t) = 0 \\ y(x,0) = y_0(x) \end{array} \right.$$

Le paramètre à identifier est $A(x)$.

On donne : $y_0(x) = \cos \pi x$

$$f(x,t) = e^{-t} \cos \pi x + \left[\frac{3}{2} - e^{-t} \right] \pi \left[\sin \pi x + (1+x)\pi \cos \pi x \right]$$

$$T = 1$$

Le choix de f a été fait de sorte que l'intégration directe du système soit possible pour une valeur particulière de $A(x)$.

On pourra ainsi faire des vérifications.

La valeur exacte de $A(x)$ qu'on veut retrouver est $A_{\text{ex}}(x) = 1 + x$

et la solution du système est

$$y(x) = \cos \pi x \left(\frac{3}{2} - e^{-t} \right)$$

L'observation est ponctuelle au point $x_0 = \frac{4,5}{20}$.

L'opérateur $\mathcal{E}$ d'observation est donc $y \xrightarrow{\mathcal{E}} \mathcal{E}y = y(x_0, \cdot)$

Donc l'observation est définie par :

$$\mathcal{E}y + \eta = y(x_0, t; A) + \eta \text{ au point courant } A.$$

Pour la résolution numérique, on va discrétiser à la fois en espace et en temps. La F.A.L. η va donc se réduire en pratique à une variable aléatoire à valeurs dans $\mathbb{R}^k$ où k est le nombre d'instantes de la discrétisation en temps.

Le problème est : $\inf_A J(A)$

$$\text{où } J(A) = \|\mathcal{E}y(A) - z_d\|^2 = \int_0^T [y(x_0, t; A) - z_d(t)]^2 dt$$

VI.1.2.- Application de l'algorithme de gradient (chap. II)

VI.1.2.1.- Calcul du gradient de la fonctionnelle

Suivant les résultats de G. Chavent [1], on introduit le système adjoint :

$$\left| \begin{array}{l} -\frac{\partial p}{\partial t} - \frac{\partial}{\partial x} \left(A(x) \frac{\partial p}{\partial x} \right) = -2 \delta(x-x_0) [y(x_0, t; A) - z_d(t)] \\ \frac{\partial p}{\partial x}(0, t) = \frac{\partial p}{\partial x}(1, t) = 0 \\ p(x, T) = 0 \end{array} \right.$$

G. Chavent montre que le gradient $\nabla J(A)$ est alors $\nabla J(A) = \int_0^T \frac{\partial y}{\partial x} \frac{\partial p}{\partial x} dt$.

Au point courant $A(x)$, le gradient $\nabla J(A)$ s'obtient donc au prix de la résolution de 2 systèmes aux dérivées partielles : le système direct et le système adjoint. On a étudié au § II.5.3 la transmission du bruit η au gradient. On dispose donc au point A d'une observation : $\nabla J(A) + \xi$.

VI.1.2.2.- Remarques sur le contexte théorique

Toutes les hypothèses faites au chap. II pour la convergence de l'algorithme d'optimisation ne sont pas satisfaites.

Notamment, les hypothèses assurant l'unicité de la solution du problème d'optimisation ne sont pas vérifiées.

Cette difficulté n'est pas due à l'aspect stochastique du problème. Dans le cadre déterministe, G. Chavent [1] a mis en évidence des exemples où il y a effectivement plusieurs solutions.

Dans les résultats numériques, nous ne chercherons donc pas à savoir si on retrouve bien A_{exact} , mais si la valeur calculée permet de rendre compte de la solution du système et en particulier de la prédire au-delà de la période d'identification.

C'est d'ailleurs en général pour obtenir de tels résultats que l'on pose des problèmes d'identification.

VI.1.2.3.- Résultats numériques

Les calculs ont été faits sur C.I.I. 10070 avec un pas de discrétisation en espace de $\frac{1}{20}$ et en temps de $\frac{1}{20}$.

Certains résultats sont représentés sur les figures (1.1) à (3.3) ci-après dont la signification est :

Pour $i = 1, 2, 3$:

figure (i.1) : opérateur exact, opérateur initial, opérateur calculé après identification

figure (i.2) : solution calculée au point d'identification

$t \in (0,1)$: période d'identification

$t \in (1,2)$: période de prédiction

figure (i.3) : solution en $t = 2$ (fin de la période de prédiction) pour $x \in (0,1)$

- . Dans le groupe de figures $i = 1$, l'écart-type de η est 10^{-2}
- . Dans le groupe de figures $i = 2$, l'écart-type de η est 5.10^{-2} ,
les autres données étant identiques à celles du groupe $i = 1$.
- . Dans le groupe de figures $i = 3$, on change la valeur A initial utilisée dans l'algorithme d'optimisation.

En effet, comme l'avait déjà remarqué G. Chavent [1] , $\nabla J(a)(x) = 0$ aux extrémités de l'intervalle $x = 0$, $x = 1$. Donc si $A \text{ initial}(1) \neq A \text{ exact}(1)$ (par exemple), la solution du problème d'optimisation ne peut coïncider avec $A \text{ exact}$. D'où un choix de $A \text{ initial}$ tel que $A \text{ initial}(0) = A \text{ exact}(0)$,
 $A \text{ initial}(1) = A \text{ exact}(1)$.

OPERATEUR $A(x)$

VARIANCE DU BRUIT : 10^{-4}

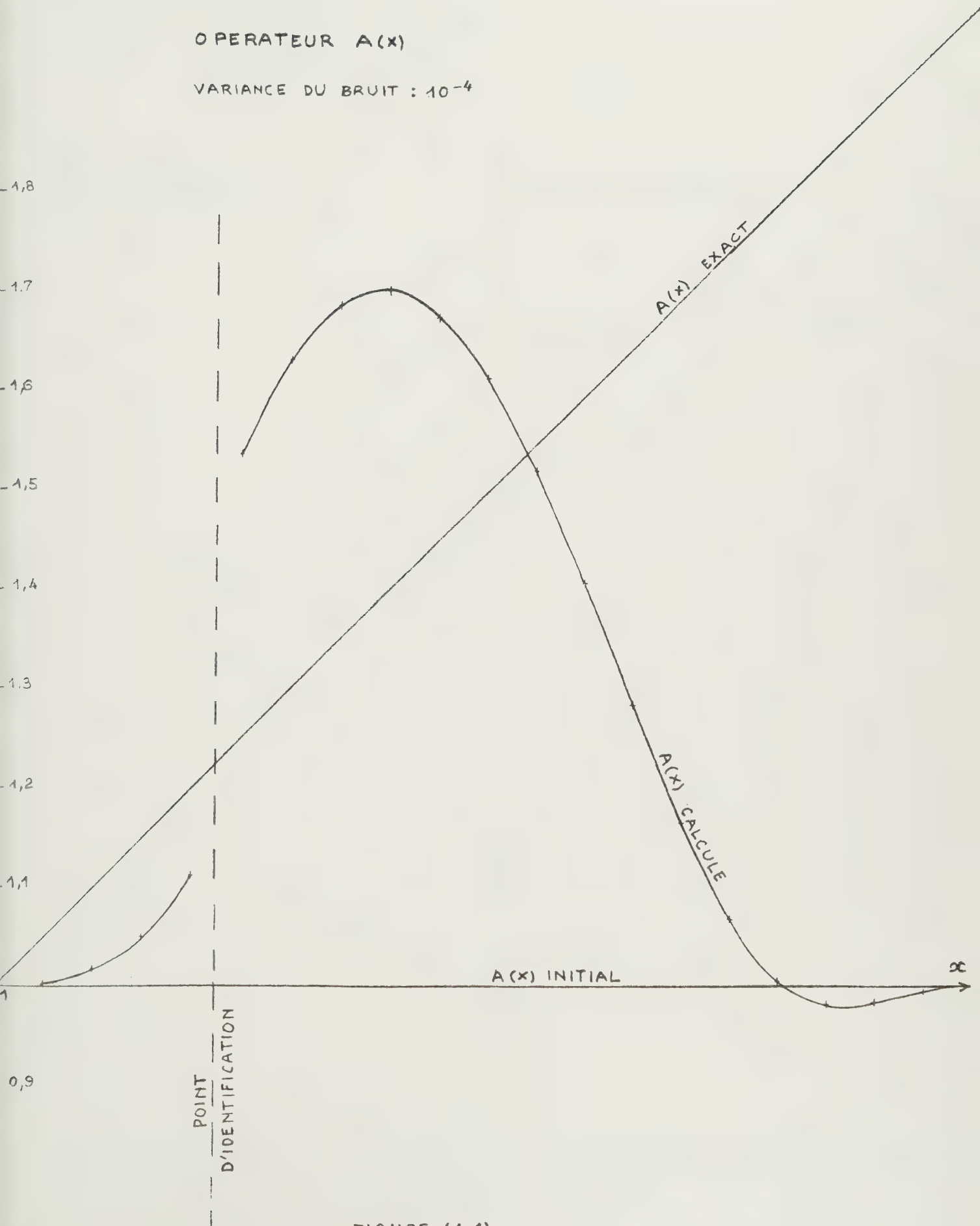
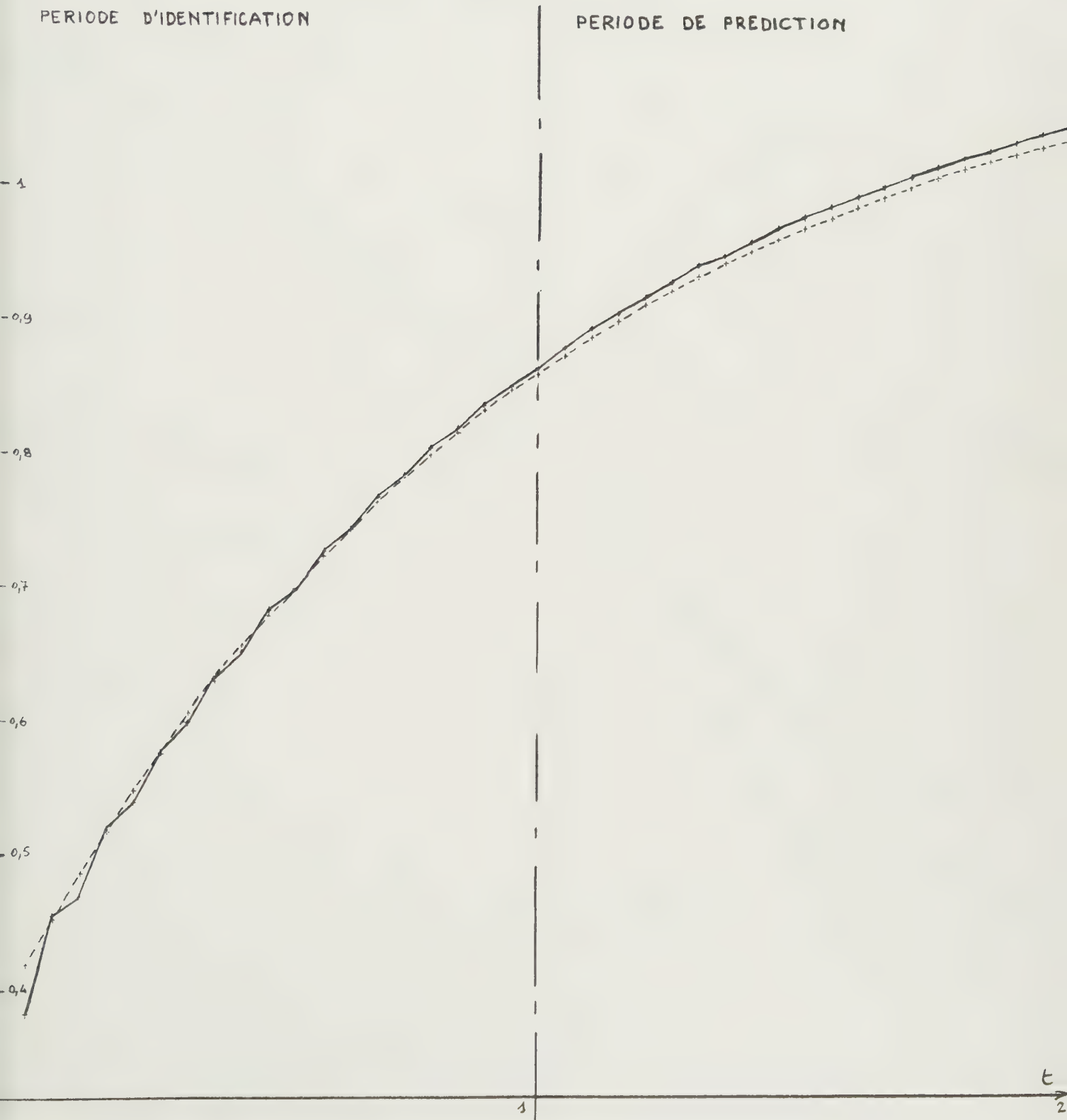


FIGURE (1,1)

SOLUTION AU POINT D'OBSERVATION

VARIANCE DU BRUIT : 10^{-4}



PREDICTION DE L'ETAT A $t=2$

VARIANCE DU BRUIT: 10^{-4}

 VALEUR EXACTE

⌈ - - - ⌋ VALEUR CALCULEE

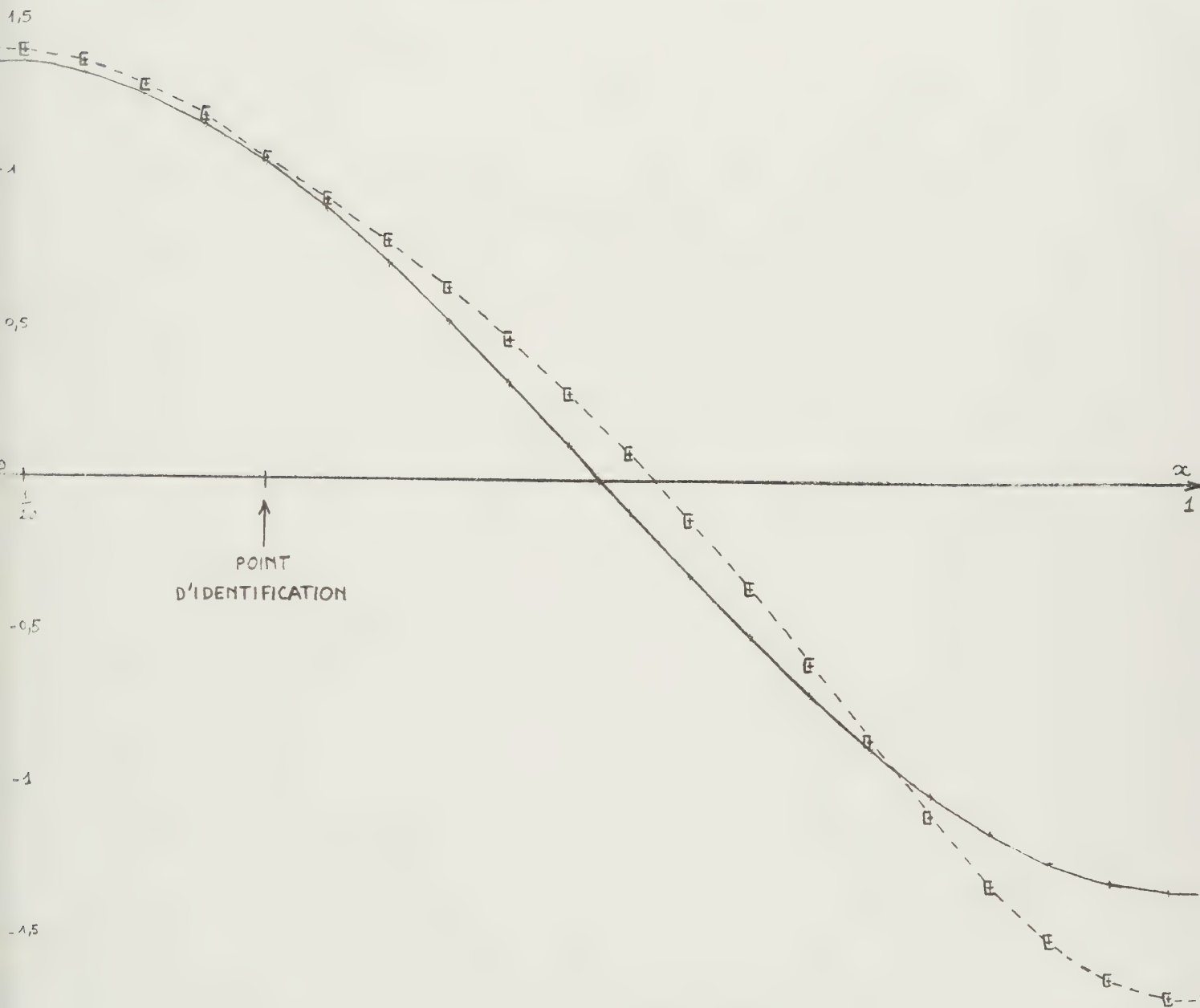


FIGURE (1.3)

OPERATEUR $A(x)$

VARIANCE DU BRUIT : $25 \cdot 10^{-4}$

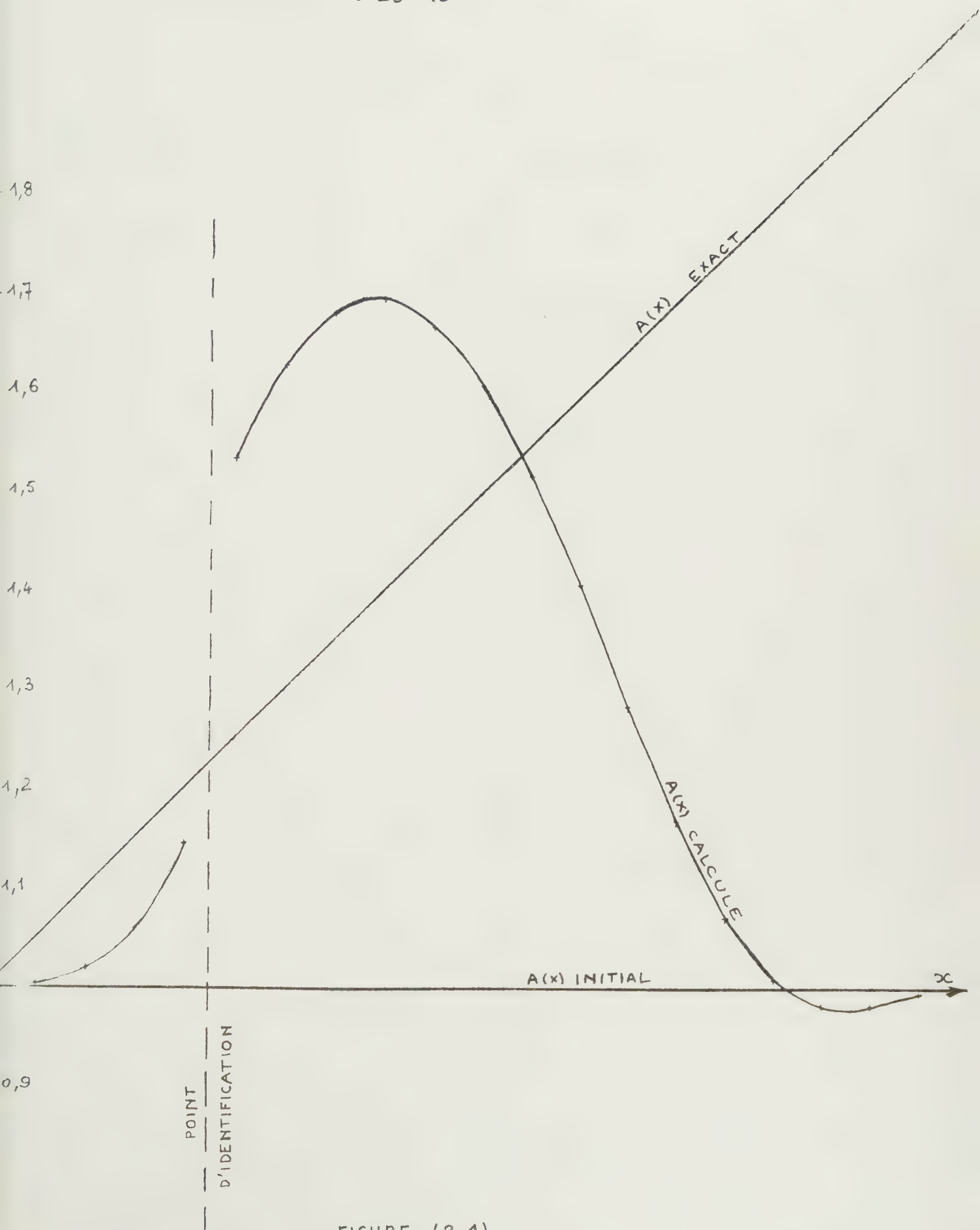


FIGURE (2,1)

SOLUTION AU POINT D'OBSERVATION

VARIANCE DU BRUIT : $25 \cdot 10^{-4}$

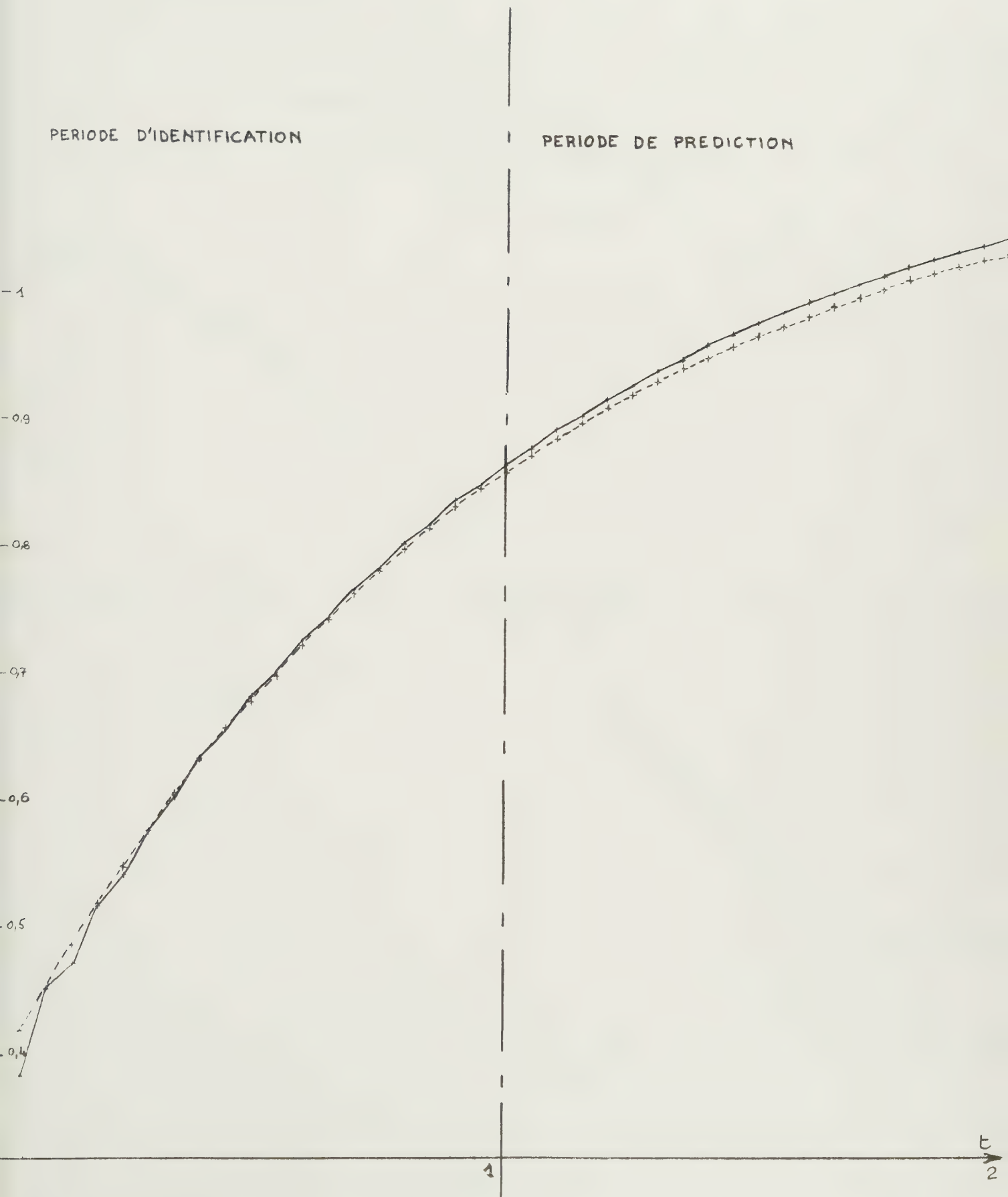


FIGURE (2,2)

PREDICTION DE L'ETAT A $t=2$

VARIANCE DU BRUIT : $25 \cdot 10^{-4}$

—+ VALEUR EXACTE

[- - - -] VALEUR CALCULEE

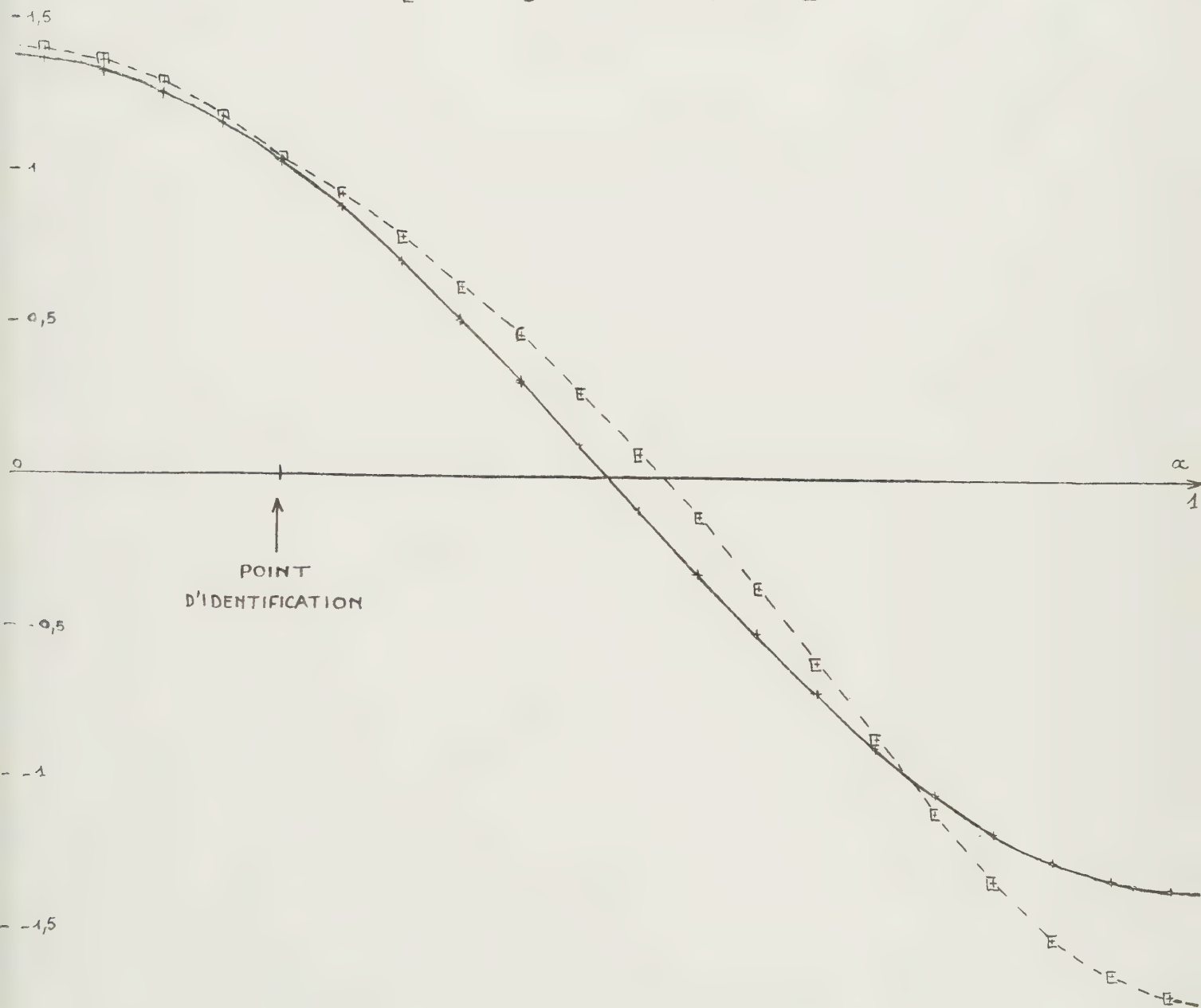


FIGURE (2,3)

OPERATEUR $A(x)$

VARIANCE DU BRUIT : $25 \cdot 10^{-4}$

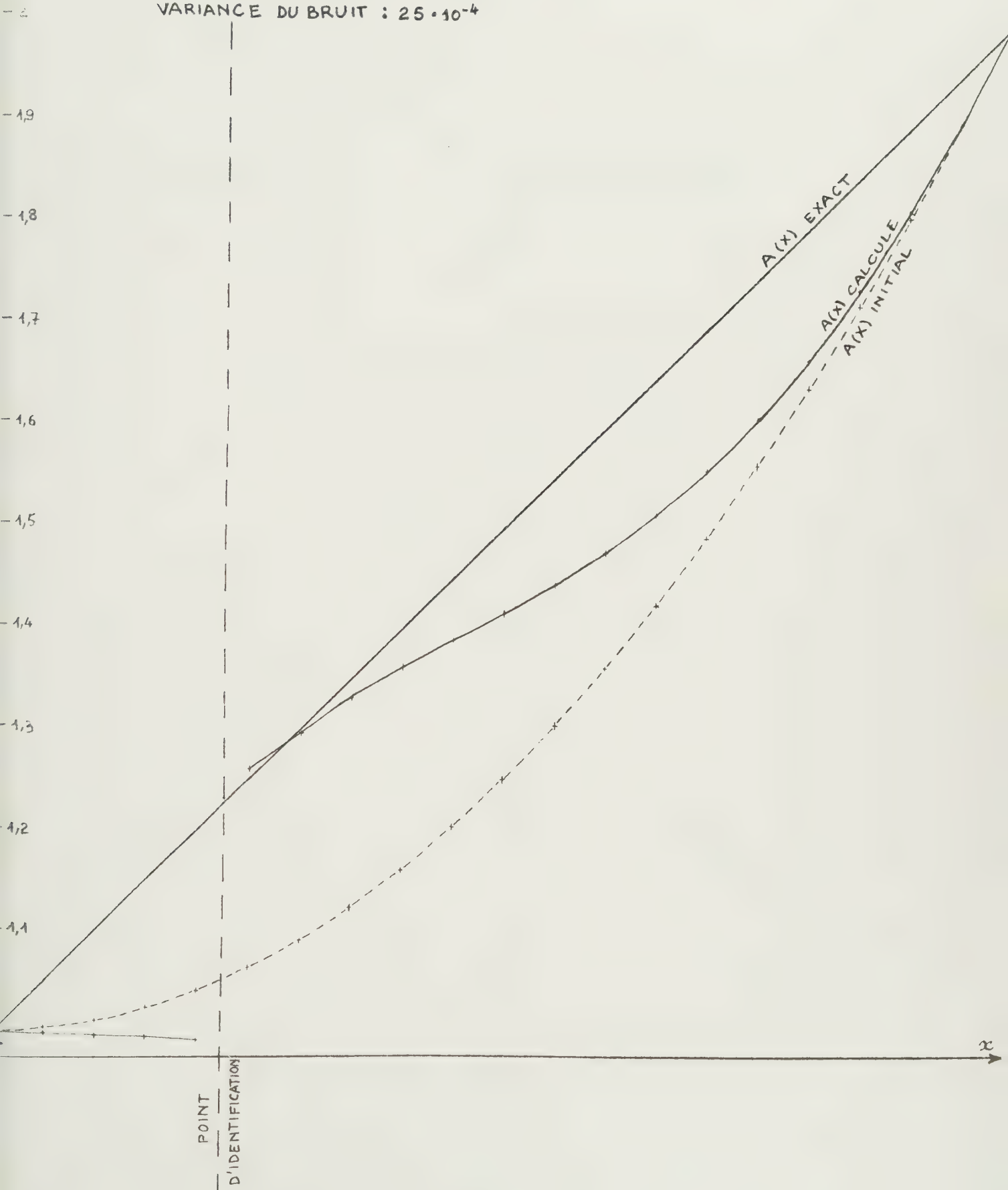


FIGURE (3,1)

SOLUTION AU POINT D'OBSERVATION

VARIANCE DU BRUIT : $25 \cdot 10^{-4}$

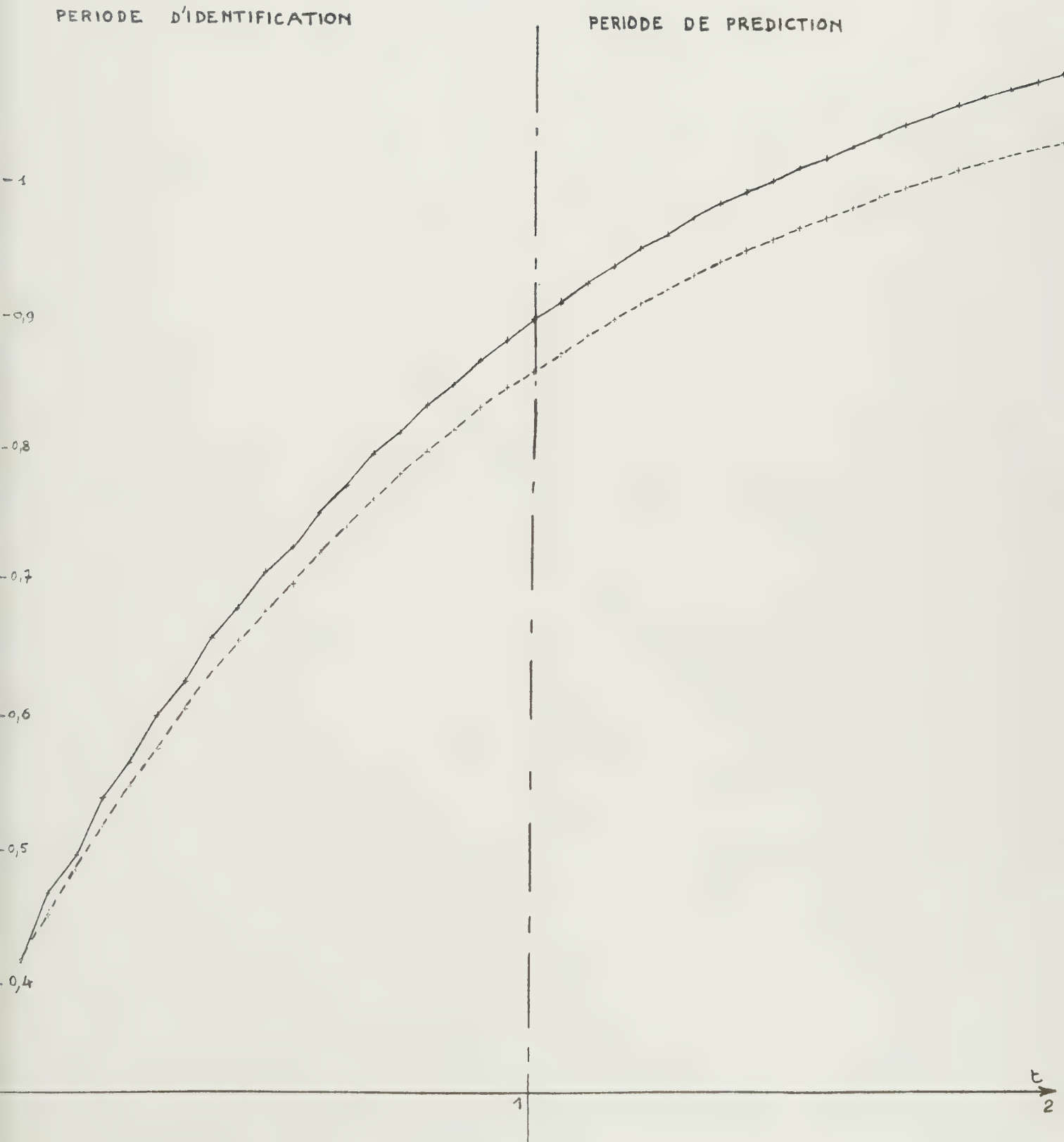


FIGURE (3,2)

PREDICTION DE L'ETAT A $t=2$

VARIANCE DU BRUIT : $25 \cdot 10^{-4}$

————— VALEUR EXACTE

- - - - - VALEUR CALCULEE

POINT
D'IDENTIFICATION

FIGURE (3,3)

Interprétation des résultats

. groupe de figures 1 :

La figure (1.1) montre qu'il n'y a pas unicité de l'opérateur vérifiant les observations.

La figure (1.2) montre qu'on obtient avec l'opérateur calculé une bonne prédiction de la solution au point d'identification.

La figure (1.3) montre qu'aux autres points, la solution calculée diffère sensiblement de la solution exacte. Avec un nombre suffisant de points d'observation ou même une observation répartie, les solutions pourraient coïncider en tout point.

. groupe de figures 2 :

Les résultats sont très proches de ceux du groupe 1.

Seule la rapidité de calcul est affectée par la multiplication par 5 de l'écart-type du bruit.

La comparaison des performances peut être résumée dans le tableau ci-dessous :

	valeur au point d'identification	valeur aux extrémités de l'intervalle		nombre d'itérations	temps de calcul	J(A) final
		$x = \frac{1}{20}$	$x = \frac{19}{20}$			
solution exacte	1,0377	1,3605	-1,3605			
solution calculée variance : 10^{-4}	1,0494	1,3988	-1,7131	3	0,066 mn	$1,045 \cdot 10^{-2}$
solution calculée variance $25 \cdot 10^{-4}$	1,0512	1,3952	-1,7122	21	0,22 mn	$3,125 \cdot 10^{-2}$

. groupe de figures 3 :

figure (3.1) : La solution se rapproche mieux de A exact. Comme aux figures (1.1) et (2.1), on constate que A calculé est tangent à A initial aux extrémités de l'intervalle et qu'il y a discontinuité de l'opérateur au point d'identification. Ces résultats (valables aussi dans le cas déterministe) s'expliquent par l'expression analytique de $VJ(A)$ (cf. G. Chavent [1]).

figures (3.2), (3.3) : Par rapport aux groupes 1 et 2, la solution est moins bonne au point d'identification. Mais elle coïncide mieux en fin de période de prédiction sur l'ensemble de l'intervalle : $x \in (0,1)$, ce qui est satisfaisant.

VI.1.3.- Application de l'algorithme de gradient approché (ch. IV)

VI.1.3.1.- Calcul du gradient approché

On discrétise le problème en espace. L'identification de $a(x)$ revient alors à la recherche d'un vecteur de $\mathbb{R}^p$ où p est le nombre de points de la discrétisation en espace : $A \approx \sum_{i=1}^p a_i \vec{u}_i$. On cherchera donc une valeur approchée de $\nabla J(A)$ par l'intermédiaire des dérivées partielles $\frac{\partial J}{\partial a_i}$, elles-mêmes approchées par différences finies. $\frac{\partial J}{\partial a_i} \approx \frac{J(A + c\vec{u}_i) - J(A)}{c}$.

Le calcul d'une valeur approchée de $\nabla J(A)$ nécessite donc la résolution de $p+1$ systèmes directs.

On voit donc que, plus la discrétisation sera fine et la dimension d'espace grande, plus la méthode sera coûteuse en temps de calcul sur ordinateur.

Les remarques sur le contexte théorique faites au § VI.1.2.2 demeurent valables.

VI.1.3.2.- Résultats numériques

Pour comparer avec le § VI.1.2.3, les données sont les mêmes.

Les résultats sont présentés sur les figures (4.1), (4.2), (4.3) qui ont même définition que les figures (1.1), (1.2), (1.3).

nombre d'itérations	temps de calcul CII 10070	J(A) final
3	0,219 mn	$1,14 \cdot 10^{-2}$

OPERATEUR $A(x)$

VARIANCE DU BRUIT : 10^{-4}

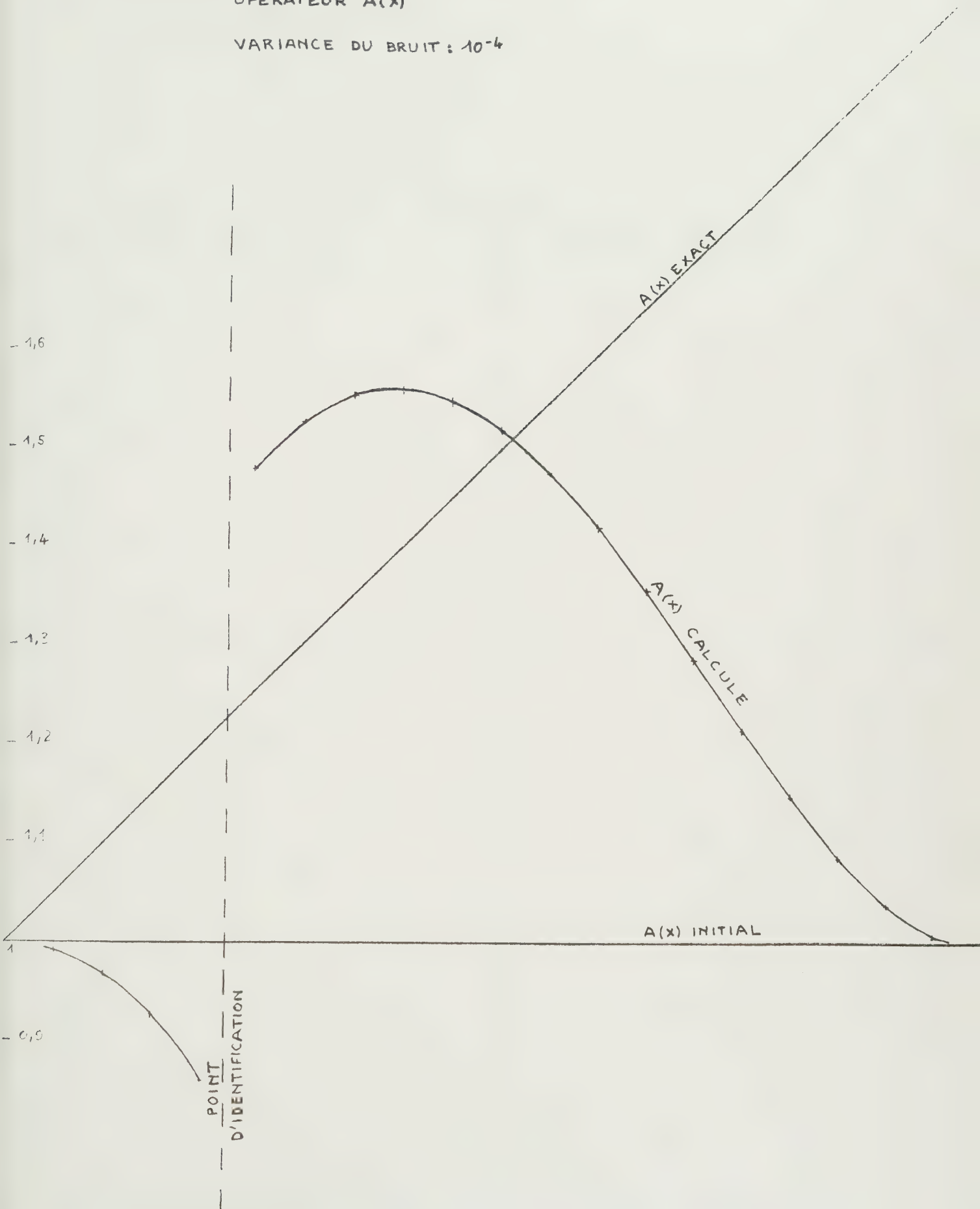


FIGURE (4,1)

SOLUTION AU POINT D'OBSERVATION

VARIANCE DU BRUIT : 10^{-4}

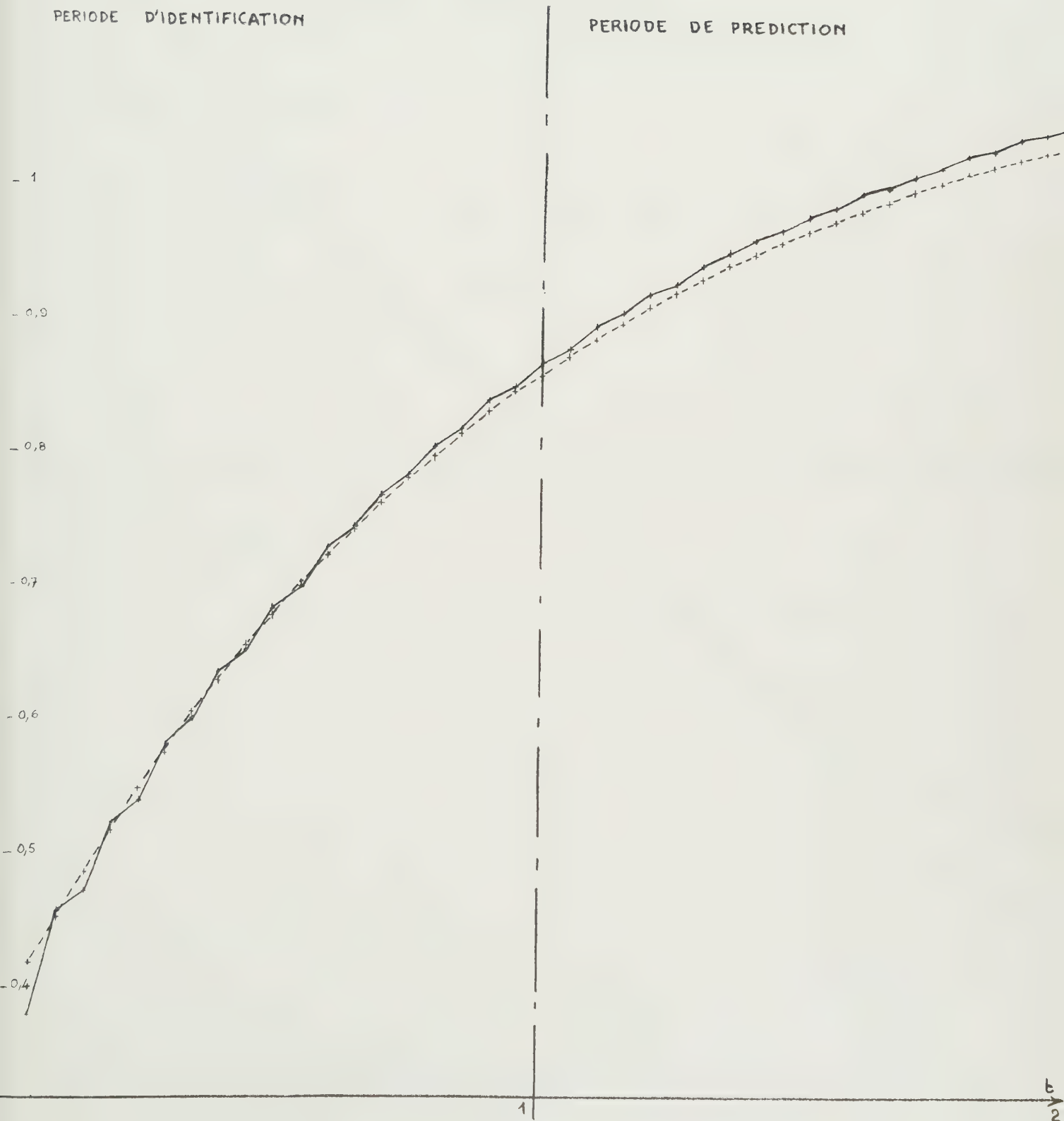


FIGURE (4,2)

Commentaires

Les commentaires du § VI.1.2.3 restent valables.

Pour une finesse de résultat comparable, le temps de calcul est nettement supérieur : il faut 5 fois plus de temps pour réaliser une itération. Cet inconvénient serait nettement aggravé si la dimension d'espace était supérieur à 1 .

Cette méthode doit donc être réservée aux cas où on ne peut accéder à la valeur stricte du gradient.

VI.2.- Un problème de lissage

VI.2.1.- Le modèle

Soit une fonction $y : x \mapsto y(x)$ de forme inconnue

$$\mathbb{R} \longrightarrow \mathbb{R}$$

On souhaite approcher $y(x)$ par une combinaison linéaire $\sum_{i=1}^k a_i \varphi_i(x)$

où les φ_i sont des fonctions de base connues.

Observation : Soient des x_j $j = 1 \dots p$ avec $p \geq k$ (pour avoir au moins autant d'informations que de grandeurs à déterminer).

Posons $\Phi = \begin{bmatrix} \varphi_1(x_1) & \dots & \varphi_k(x_1) \\ \vdots & & \vdots \\ \varphi_1(x_p) & \dots & \varphi_k(x_p) \end{bmatrix}$

On suppose que les φ_i sont telles qu'on puisse trouver des x_j tels que Φ soit de rang k ($\leq p$) .

Ces x_j une fois choisis, on sait observer des réalisations de la variable aléatoire $Y + \eta(x)$ à valeurs dans $\mathbb{R}^p$

où $Y = \begin{bmatrix} y(x_1) \\ \vdots \\ y(x_p) \end{bmatrix}$ $\eta(x) = \begin{bmatrix} \eta_1(x_1) \\ \vdots \\ \eta_p(x_p) \end{bmatrix}$ $E(\eta(x)) = 0$.

Le problème est de choisir $A = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}$

de façon à minimiser $f(A) = \|Y - \Phi A\|^2 = \sum_{j=1}^p \left[y(x_j) - \sum_{i=1}^k a_i \varphi_i(x_j) \right]^2$

au vu de l'observation $Y + \eta(x)$.

$$\nabla f(A) = -2 {}^t\Phi [Y - \Phi A] .$$

Pour A fixé quelconque, en ajoutant $-\Phi A$ au résultat d'une observation (expérience) et en multipliant par $-2 {}^t\Phi$, on obtient une réalisation de la variable aléatoire $\nabla f(A) + Z_X$ où $Z_X = -2 {}^t\Phi \eta(x)$. Z_X ne dépend pas de A et $E(Z_X) = 0$.

VI.2.2.- Application des théorèmes d'optimisation stochastique

Les hypothèses déterministes faites au chap. II sur la fonctionnelle f sont vérifiées.

En effet, ${}^t\Phi\Phi$, matrice $k \times k$ symétrique, est définie positive car elle est de rang k , ce qui résulte de :

$${}^t\Phi(\mathbb{R}^p) = {}^t\Phi(\Phi(\mathbb{R}^k)) \oplus {}^t\Phi(\Phi(\mathbb{R}^k)^\perp)$$

$$\text{et de } y \in \Phi(\mathbb{R}^k)^\perp \iff y \in \text{Ker}({}^t\Phi) .$$

De plus $\Phi {}^t\Phi$ est de rang k , ce qui résulte de l'inégalité de Sylvester :
 $r(A) + r(B) - n \leq \text{rang}(AB) \leq \inf(r(A), r(B))$ où n = nombre de colonnes de A = nombre de lignes de B .

L'hypothèse H'_6 se traduirait par (cas sans contrainte)

$$E(\|Z_X\|^2) \leq B \|\nabla f(A)\|^2 .$$

Elle ne sera pas vérifiée dans un exemple réel provenant de vraies mesures.

Cependant, pour tester numériquement le théorème 2.3.2.2, on peut simuler une telle observation :

1- Pour pouvoir contrôler les résultats, on se donne à l'isser une fonction $y(x)$ qui est justement de la forme $y(x) = \sum_{i=1}^k a_i \psi_i(x)$ et on se donne les a_i exacts.

2- Pour A donné (différent de A exact), on peut calculer
 $-2 {}^t\Phi [Y - \Phi A] = \nabla f(A) .$

3- On génère par l'ordinateur des réalisations de variables aléatoires $\xi_1 \dots \xi_k$ telles que $E(\xi_i) = 0$, $E(\xi_i^2) \leq B$.

L'observation simulée sera $\begin{bmatrix} 1+\xi_1 & \dots & 1+\xi_k \end{bmatrix} \nabla f(A)$

de sorte que H'_6 soit vérifiée.

L'hypothèse $\underline{H_6}$, suffisante pour appliquer le théorème 2.2.2.2 et le théorème 2.3.2.4 sera elle vérifiée dans les exemples réels sous l'hypothèse raisonnable : $E \left[(\eta_i(x_i))^2 \right] \leq c$.

VI.2.3.- Exemple numérique

Soit $\dot{x}_i(x) = x^{i-1}$. On veut donc faire un lissage par un polynôme, et pour pouvoir contrôler les résultats, on choisit pour $y(x)$ un polynôme de degré $k-1$.

$$\Phi = \begin{bmatrix} 1 & x_1 & \dots & x_1^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_p & \dots & x_p^{k-1} \end{bmatrix} \text{ est de rang } k \text{ (matrice de Van der Monde) .}$$

On a pris $p = k = 3$

$$x_1 = -1, \quad x_2 = 0, \quad x_3 = +1 \quad \text{et} \quad y(x) = 1 + x + x^2.$$

Quelques résultats avec observation simulée (vérifiant $\underline{H'_6}$) sont présentés sur les fig. (5.1) : pas constant

fig. (5.2) : pas optimisé à chaque itération.

Le choix judicieux du pas améliore nettement la convergence.

Lorsque seule $\underline{H_6}$ est vérifiée, on a pris $B^2 = 10^{-2}$. Alors $\frac{B^2}{\Lambda^2} \approx 10^{-4}$ on constate effectivement une marche aléatoire dans une boule de rayon 10^{-2} avec quelques sorties (on y reste en moyenne quadratique !).

Lors des sorties de la boule, l'algorithme ramène rapidement le processus dans la boule.

Lorsqu'on utilise un algorithme mixte, c'est-à-dire qu'on remplace le pas par le terme général d'une série divergente une fois que le 1er algorithme est entré dans la boule, on constate une convergence très lente vers la solution.

Deux causes de lenteur sont cumulées :

- méthode à petit pas
- bruit aléatoire.

L'application dès le départ de l'algorithme de série divergente dès les premières itérations rendrait ces difficultés sensibles dès le début alors que la méthode mixte permet de s'approcher d'abord à grands pas de la solution.

PAS CONSTANT

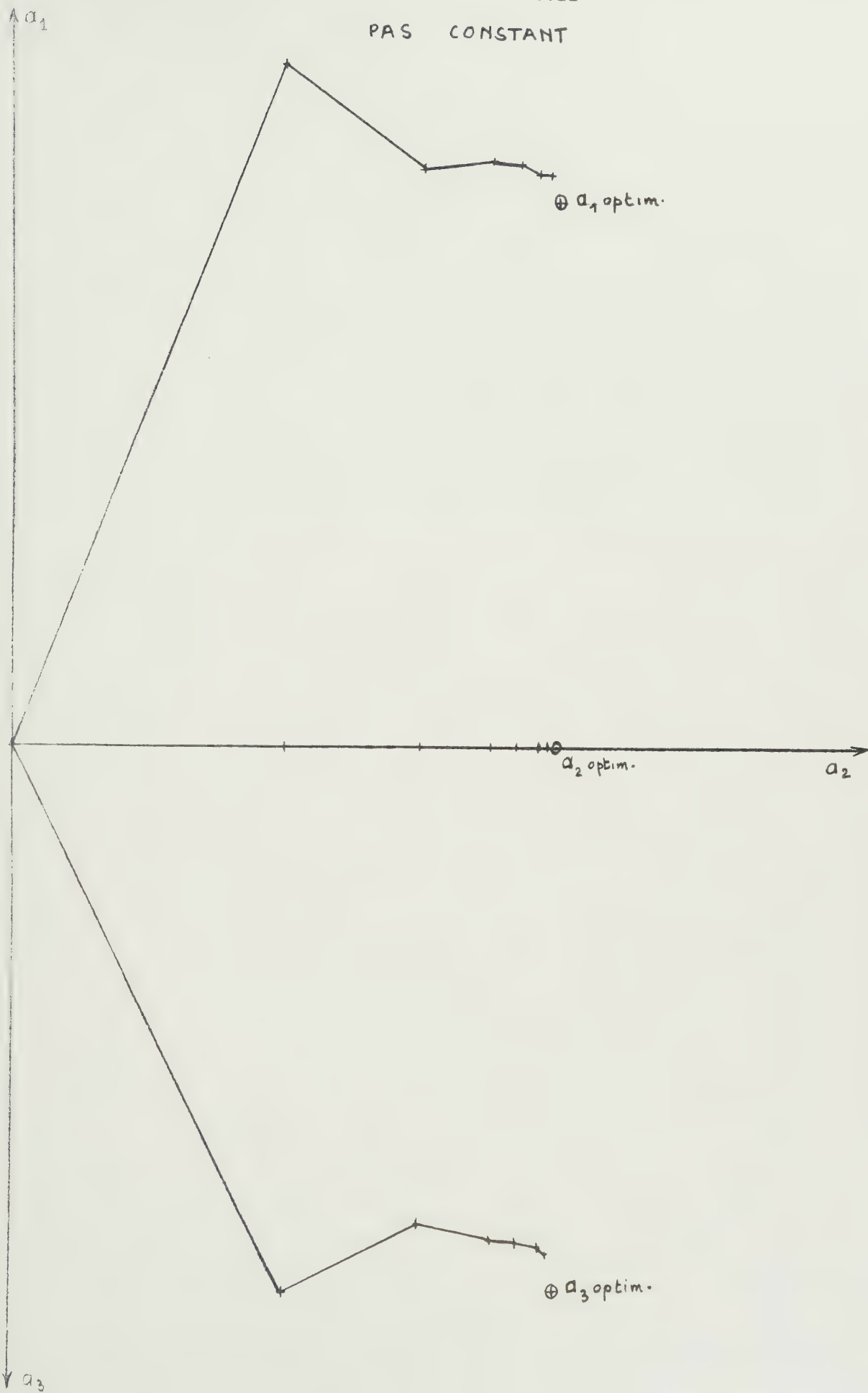


FIGURE (5,1)

PAS OPTIMAL

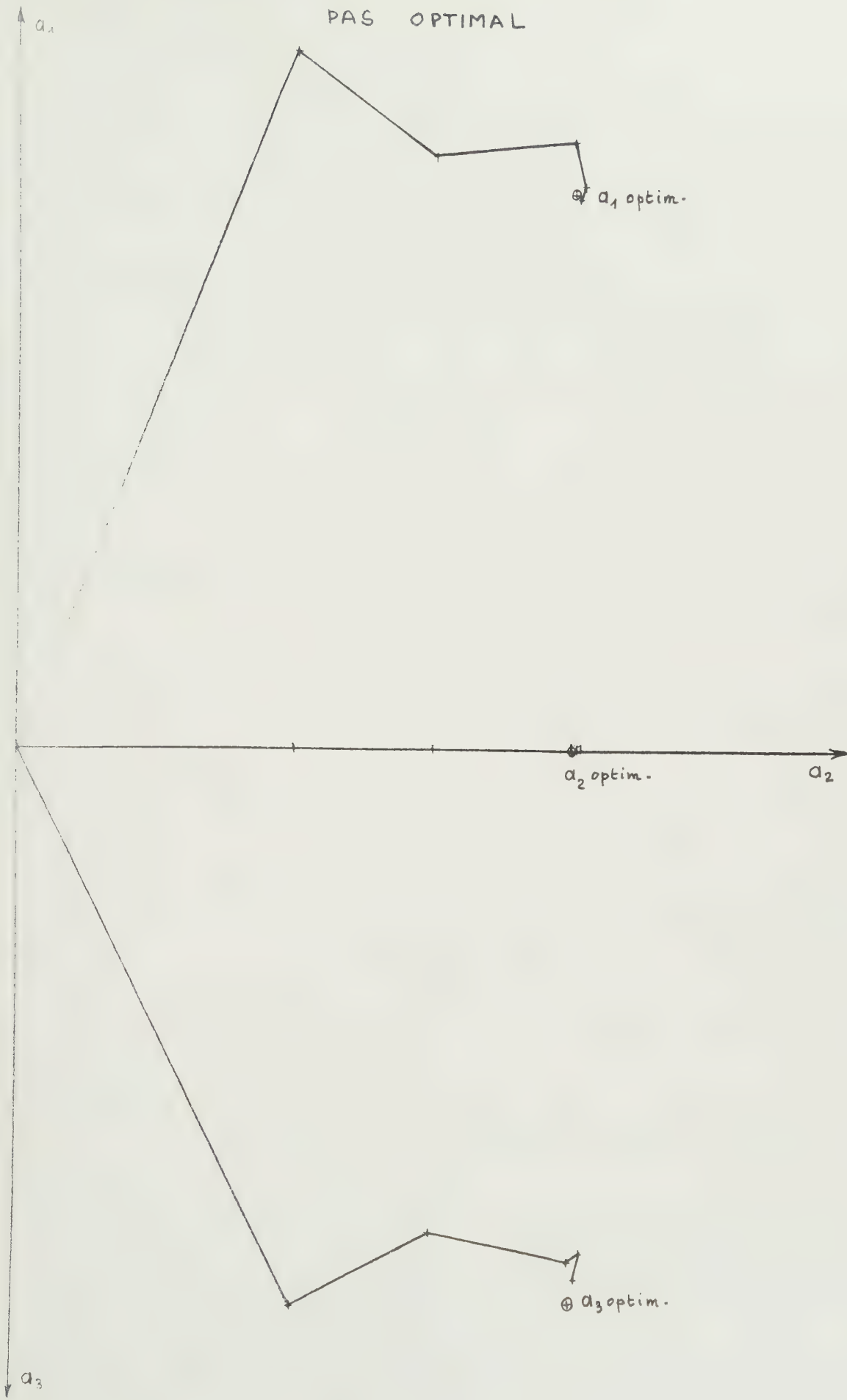


FIGURE (5,2)

Bibliographie

- K.J. Arrow, L. Hurwicz, H. Uzawa [1] Studies in linear and non linear programming, Stanford University Press (1958)
- A. Auslender [1] Méthodes et théorèmes de dualité
RIRO (4^e année, n° R1, 1970, pp. 9-45)
- [2] Problèmes de Minimax via l'analyse convexe et les inégalités variationnelles : théorie et Algorithmes
Springer (1972)
- A. Bensoussan [1] Filtrage optimal des systèmes linéaires, Dunod (1971)
- [2] L'Identification et le filtrage, cahier IRIA n° 1 (Fév. 1969)
- A. Bensoussan, J.L. Lions, R. Tremam [1] Sur les méthodes de décomposition, de décentralisation et de coordination et applications, cahier IRIA n° 11 (Juin 1972)
- J.P. Bertran [1] Optimisation stochastique dans un espace de Hilbert, Méthode de gradient, C.R. Ac. Sc. Paris t. 276, Série A, pp. 613-616 (1973)
- [2] Optimisation stochastique dans un espace de Hilbert, Colloque Analyse numérique, Epinal (1972)
- [3] Optimisation stochastique dans un espace de Hilbert, méthode de séries divergentes, Colloque Analyse numérique, La Colle/Loup (1973)
- [4] Méthodes stochastiques de dualité, Colloque Analyse numérique, Gourette (1974)
- J. Bourbaki [1] livre VI - Intégration, chap. 1.4 , Hermann A.S.I. 1175
- [2] livre V - Espaces vectoriels topologiques, chap. 3.5, Hermann A.S.I. 1220
- J. Cea [1] Optimisation - Théorie et Algorithmes, Dunod (1971)
- J. Cea, R. Glowinski [1] Minimisation de fonctionnelles non différentiables, IRIA IRIE/7105 (mars 1971)
- G. Chavent [1] Analyse fonctionnelle et identification de coefficients répartis dans les équations aux dérivées partielles, Thèse Paris (1971)

- G.P. Comer [1] Some stochastic approximation procedures for use in Process control - Ann. Math. Stat. vol. 35, n° 3, pp. 1136-1146 (1964)
- E.L. Doob [1] Stochastic processes - Wiley New-York (1953)
- E.E. Dubbins, D.A. Friedman [1] A sharper form of the Borel-Cantelli lemma and the strong law - Ann. Math. Stat. vol. 36, pp. 800-807 (1965)
- H. Engelhardt [1] On upper bounds for variances in stochastic approximation SIAM J. Appl. Math. vol. 24, n° 2, pp. 145-151 (1973)
- G. Fichet [1] Sur l'approximation et l'optimisation stochastique - Thèse d'ingénieur docteur, Toulouse (1970)
- L. Gladyshev [1] On stochastic approximation - Theory of Probability and its applications - vol. X, n° 2, pp. 275-278 (1965)
- A.A. Goldstein [1] Minimizing functionals on Hilbert spaces - computing Methods in optimization problems - Academic Press (1964)
- J. Hans, M. Driml [1] Conditionnal expectations for generalized random variables - Trans. second Prague conf. on Info. Theory - Prague (1960)
- P. Kenneth, M. Silony, J.P. Yvon [1] La méthode de pénalisation et ses applications aux problèmes de contrôle optimal - Cahier IRIA n° 2 - Algorithmes numériques d'optimisation (1970)
- J. Kiefer, J. Wolfowitz [1] Stochastic estimation of the maximum of a regression function, Ann. Math. Stat. vol. 33, pp. 462-466 (1952)
- H.J. Kushner [1] Stochastic approximation algorithms for the local optimization of functions with nonunique stationary points - IEEE Trans. Autom. Contr. vol. 17, pp. 646-654 (1972)
- H.J. Kushner [2] Stochastic approximation algorithms for constrained optimization problems - Brown University, Providence R.I. CDS Rep. 72.1 et à paraître Ann. Statist.
- H.J. Kushner, T. Gavin [1] Stochastic approximation type methods for constrained systems : Algorithms and numerical results - IEEE Trans. Autom. Contr. vol. AC 19, pp. 349-357 (1974)

- J.L. Lions [1] Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles - Dunod (1968)
- J.L. Lions, R. Trémoilières, R. Glowinski [1] Livre à paraître sur l'Analyse numérique des Inéquations variationnelles - Dunod
- B. Martinet [1] Maximisation d'une fonction concave connue de façon approchée - Application à la méthode duale de Rockafellar - Colloque d'analyse numérique, la Colle/Loup (1973)
- M. Métivier [1] Martingales à valeurs vectorielles, applications à la dérivation des mesures vectorielles - Ann. Inst. Fourier, Grenoble 17.2 (1967) pp. 175-208
- E. Mourier [1] Eléments aléatoires dans un espace de Banach - Thèse Paris (1952)
- J. Neveu [1] Calcul des probabilités - Masson - Paris (1970)
- B.T. Poljak [1] A general method of solving extremum problems - Soviet Math. Dokl. vol. 8, n° 3, pp. 593-597 (1967)
- P. Révész [1] Robbins-Monro procedure in a Hilbert space and its applications in the theory of learning processes (I et II) - Studia Scientiarum Mathematicarum Hungarica 8 (1973) pp. 391-398 et pp. 469-472
- H. Robbins, S. Monro [1] A stochastic approximation method - Ann. Math. Stat. vol. 22, pp. 400-407 (1951)
- R.T. Rockafellar [1] A dual approach to solving non linear programming problems by unconstrained optimization - Mathematical programming pp. 354-473 (1973)
- Jocelyne Rouyer [1] Propriétés stochastiques du gradient de la fonctionnelle coût dans la méthode de l'état adjoint lorsque l'observation est stochastique - Rapport de DEA - Nancy (1972)
- Joseph Rouyer [1] Méthodes de Frank-Wolfe appliquée à l'optimisation stochastique Rapport de DEA - Nancy (1973)
- J. Sacks [1] Asymptotic distributions of stochastic approximation procedures - Ann. Math. Stat. vol. 29, pp. 373-405 (1958)
- L. Schemetterer [1] Sur l'itération stochastique - Le calcul des probabilités et ses applications 87 pp. 55-63 (1958)

- L. Schemetterer [2] Multidimensional stochastic approximation - Multivariate Analysis II - Proc. 2nd Int. Symp. Dayton Ohio - Ac. Press. New-York, pp. 443-460 (1969)
- J.H. Venter [1] On Dvoretzky stochastic approximation theorems - Ann. Math. Stat. vol. 37, pp. 1534-1544 (1966)
- [2] On convergence of the Kiefer-Wolfowitz approximation procedures - Ann. Math. Stat. vol. 38, pp. 1031-1036 (1967)

NOM DE L'ETUDIANT : BERTRAN Jean Pierre

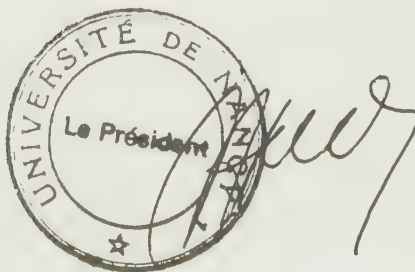
NATURE DE LA THESE : DOCTORAT D'ETAT ES SCIENCES

VU, APPROUVE

& PERMIS D'IMPRIMER

NANCY, le 13 Mai 1975

LE PRESIDENT DE L'UNIVERSITE DE NANCY I



J.R. HELLOY

